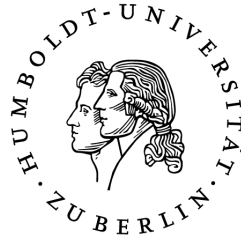


HUMBOLDT-UNIVERSITÄT ZU BERLIN

INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN

ZUR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT

HEFT 357

OA_SCORE – ÜBERLEGUNGEN ZU EINER METHODE DER
QUALITÄTSBESTIMMUNG ELEKTRONISCHER VERÖFFENTLICHUNGEN MITTELS
STANDARDISIERTER BEWERTUNG DURCH DIE LESER.

VON

UWE JUNG

OA_SCORE – ÜBERLEGUNGEN ZU EINER METHODE DER
QUALITÄTSBESTIMMUNG ELEKTRONISCHER VERÖFFENTLICHUNGEN MITTELS
STANDARDISIERTER BEWERTUNG DURCH DIE LESER.

VON
UWE JUNG

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 357

Jung, Uwe

OA_SCORE – Überlegungen zu einer Methode der Qualitätsbestimmung elektronischer Veröffentlichungen mittels standardisierter Bewertung durch die Leser. / von Uwe Jung. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2014. - 63 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 357)

ISSN 14 38-76 62

Abstract:

Nicht bzw. nicht ausreichend begutachtete Veröffentlichungen spielen in der Diskussion um die Qualitätsmessung wissenschaftlicher Arbeiten bisher nur eine untergeordnete Rolle. Angesichts der aktuellen Entwicklung von Open Access steht jedoch zu vermuten, dass deren Zahl und wohl auch deren Bedeutung aus verschiedenen Gründen stark zunehmen wird.

In dieser Arbeit wird das Grundkonzept eines Messinstruments entwickelt, welches in der Lage sein soll, Hinweise auf die Qualität einer Veröffentlichung zu geben. Dieses hier OA_SCORE genannte Instrument basiert auf einer standardisierten Leser-Rückmeldung in Form einer in das Dokument integrierten multidimensionalen Ein-Klick-Lösung.

OA_SCORE versteht sich als Teil des Altmetrics-Ansatzes. Der Anwendungsbereich kann auch auf andere Arten wissenschaftlicher Veröffentlichungen ausgedehnt werden.

Dieser Text geht zurück auf eine Masterarbeit im postgradualen Fernstudiengang M. A. Bibliotheks- und Informationswissenschaft (Library and Information Science) an der Humboldt-Universität zu Berlin.

Online-Version: <http://edoc.hu-berlin.de/series/berliner-handreichungen/2014-357>



Dieses Werk steht unter einer Creative Commons **Namensnennung-NichtKommerziell-KeineBearbeitung 3.0 Deutschland-Lizenz**.

Die Arbeit wurde in Jaunde / Republik Kamerun erstellt.

Inhaltsverzeichnis

1	Einleitung	7
2	Qualität und Qualitätsfilter	8
3	Ausgangsidee	13
4	Forschungsfragen	17
5	Methoden	17
5.1	Technische Anwendbarkeit	17
5.2	Bedarfsermittlung	20
5.3	Plausibilität der gewonnenen Aussagen	21
5.3.1	Bewertungsdesign	21
5.3.2	Pretest	29
6	Untersuchungen	31
6.1	Metadaten-Abfrage von OA-Repositorien	31
6.1.1	Datenkorpus	31
6.1.2	Ergebnisse	36
6.2	Bewertungsdesign	42
6.2.1	Datenkorpus	42
6.2.2	Ergebnisse	43
6.2.3	Zusammenfassung zum Bewertungsdesign	48
7	Weitere Überlegungen zum Design	49
7.1	Darstellung der Skala	49
7.2	Implementierung	50
7.3	Motivation der Teilnehmer	53
8	Schluss und Ausblick	55
9	Literaturverzeichnis	57
9.1	Institutionelle URL	57
9.2	Publikationen	57
10	Anhang	61
10.1	Häufigkeits- und Synonymanalyse der Begriffsgruppen Originalität, Plausibilität und Nützlichkeit	61
10.2	Rohdaten der Meta-Repositorien	61
10.2.1	OpenDOAR	61
10.2.2	ROAR	61
10.3	Konsolenbefehle zur Abfrage der Meta-Repositorien mittels xmlstart	61
10.4	BASH-Script zur Datenabfrage nach OAI-PMH	61
10.5	Verwendete Fragebögen	61
10.6	Korrigierte Rohdaten der Repositorien	62
10.7	Analyse der Metadatenformate	62
10.8	Untersuchung zur Auftreten von Formatangaben	62
10.9	Überprüfung zur Zahl der begutachteten Items	62
10.10	Ergebnisse der Befragung	62
10.11	XML-Schema	62

Hinweis: Der Anhang besteht aus insgesamt 23 Dateien. Sie sind in geeigneter Form dieser Arbeit beigelegt.

1 Einleitung

Kein Zweifel, die Open-Access-Bewegung gewinnt an Zulauf. Ihre öffentliche Wahrnehmung nimmt seit Jahren zu. Viel beachtete Erklärungen von Universitäten und Forschungsfinanzierern zugunsten von Open Access (OA) verfehlen ihre Wirkung nicht.

Im Mai 2012 waren 35,79 Millionen frei zugängliche Veröffentlichungen bei der *Bielefeld Academic Search Engine* (**BASE**) registriert (**Vierkant 2012, 55**). Im Mai 2013 konnte man in der Elektronischen Zeitschriftenbibliothek (**EZB**) mehr als 40.000 Einträge zu frei zugänglichen Zeitschriften finden. Das *Open Directory of Open Access Repositories* (**Open-DOAR**) verzeichnete im März 2012 insgesamt 2.212 Einträge. Und am 12. März 2013 vermeldete das *Directory of Open Access Journals* (**DOAJ**) die Indexierung des millionsten Artikels. Auch wenn darunter Mehrfachnennungen sind, so geben die Zahlen doch einen Eindruck von der Menge der frei verfügbaren wissenschaftlichen Literatur im Netz.

Das wirft die Frage auf, mit welchen Methoden sich diese Veröffentlichungen am besten filtern lassen. Die Diskussion um das Für und Wider bestimmter Verfahren ist älter als der Begriff der Bibliometrie (**Jovanovic 2012, 73**). Seitdem jedoch das *World Wide Web* immer stärker als Veröffentlichungsplattform genutzt wird, bieten sich zunehmend Alternativen an.

Unter einem Filter in der hier verwendeten Form soll ein Instrument verstanden werden, welches Aussagen über die Qualität eines wissenschaftlichen Beitrages ermöglicht. Diese Aussagen wiederum sollen dem Leser dabei helfen, die knappe Ressource Zeit optimal einzusetzen, indem er nur solche Beiträge vertiefend zur Kenntnis nimmt, die für seine Arbeit innovativ, plausibel und relevant sind.

„Traditionelle“ Filter, wie z. B. Peer-Review-Verfahren oder Zitationsanalysen, sind wegen ihrer praktischen Anwendung in die Kritik geraten. An das Medium Internet gebundene Filter gewinnen an Bedeutung, lassen aber noch methodische Defizite erkennen.

Es ist abzusehen, dass zukünftig ein Mix aus verschiedenen Filtern und Indikatoren für Aussagen zur Qualität von wissenschaftlichen Veröffentlichungen herangezogen werden wird. Diese Arbeit möchte einen neuen Filter vorstellen und im Hinblick auf seine praktische Anwendbarkeit hinterfragen.

Der wichtigste mögliche Anwendungsbereich dieses neuen Instruments sind nicht-begutachtete Open-Access-Veröffentlichungen. Sie stellen eine Untermenge der sogenannten *Green-Road-Veröffentlichungen* dar. Diese Gruppe von Veröffentlichungen spielt bei der Diskussion von qualitativen Filtern bisher nur eine untergeordnete Rolle. (**Oppenheim 2008, 580**). Das erscheint angesichts der bis dato vorherrschenden Kultur wissenschaftlicher Veröffentlichungen durchaus der Norm zu entsprechen. Begutachtete Zeitschriftenartikel bzw. Qualitätsarbeiten sind darin vorherrschend. Darüber hinaus existieren aber auch Preprints, Workshopbeiträge, Primärdaten, Discussion Papers, usw., bei denen Begutachtung keine bzw. nur geringe Anwendung findet.

Aber auch durch die Kritik an den Methoden der Begutachtung wird deutlich, dass deren Ergebnisse immer auch ein Kontinuum darstellen. Es gibt keine plausible Einteilung von Veröffentlichungen einerseits in qualitativ abgesicherte – weil begutachtete – und andererseits mit Vorsicht zu genießenden nicht begutachteten Publikationen. Die Realität lehrt uns, dass in Bezug auf die Qualität die Übergänge fließend sind.

Es gibt einen weiteren Grund, warum nicht-begutachteten OA-Veröffentlichungen zukünftig mehr Beachtung geschenkt werden sollte. Eines der häufig wiederkehrenden Argumente für Open Access ist die damit einhergehende Verfügbarkeit von wissenschaftlichen Informationen in Ländern der sogenannten Dritten Welt. Diese argumentative Verbindung von Open

Access und Entwicklungsdiskurs hat im Verlaufe der Jahre stetig zugenommen (**Haider 2012, 72 ff**).

Es stimmt, dass der Zugriff auf kostenpflichtige Angebote der großen Wissenschaftsverlage von diesen Ländern aus kaum realisiert ist. Es wäre aber m. E. zu kurz gegriffen, hierfür nur den Mangel an finanziellen Ressourcen als Ursache zu benennen. In nicht wenigen der betroffenen Gesellschaften stimmt die vorherrschende wissenschaftliche Kultur nicht mit derjenigen Auffassung überein, wie sie für einen globalen Wissenschaftsbetrieb geboten erscheint.

Wie ist es zum Beispiel zu erklären, dass die Zahl der in afrikanischen Ländern beheimateten OA-Repositoryn im internationalen Maßstab kaum wahrnehmbar ist? Gleichwohl sollten diese doch der Theorie nach eine anzustrebende Alternative zu kostenpflichtigen Angeboten darstellen. Die über **OpenDOAR** oder **Ranking Web of Universities** abrufbaren Statistiken sind hier eindeutig. Besonders augenfällig wird dieses Phänomen in den frankophonen Ländern der Region. Ein wichtiger Grund hierfür ist wahrscheinlich darin zu suchen, dass Wissenschaftsbetrieb hier „anders“ funktioniert, als in Europa oder Nordamerika. Forschungsförderung auf nationaler Ebene ist in diesen Ländern praktisch nicht existent. Etablierte Führungskräfte in diesem System haben wenig Möglichkeiten und auch Interesse an einer persönlichen Weiterentwicklung. Mitunter nutzen sie ihre Verfügungsgewalt über die wenigen vorhandenen Ressourcen dazu, nachwachsende Konkurrenten vom Zugriff auf eben diese Ressourcen fernzuhalten und die eigene Klientel zu favorisieren. Im Ergebnis bieten sich diesen Nachwuchswissenschaftlern kaum Möglichkeiten, vor Ort angemessen zu publizieren. Die Zahl der Zeitschriften ist gering, viele davon erscheinen nach wie vor nur als Printausgabe und werden daher in der übrigen Welt kaum rezipiert. Wenige herausragende Vertreter ihres Faches haben die Möglichkeit, über ausländische Förderprogramme in renommierten Journalen zu veröffentlichen. Für die Mehrzahl der an einer Karriere interessierten Wissenschaftler hat sich hingegen extra ein eigener Markt von dubiosen Online-Verlagen gebildet. Nicht zufällig residieren einige bedeutende dieser als *Open Access Predators* bekannten Unternehmungen in afrikanischen Ländern. In der 2012er Ausgabe von Jeffrey Bealls' **List of Predatory Open Access Publishers** werden drei dieser Verlage auf einen nigerianischen Ursprung zurückgeführt, darunter das den Bereich stark dominierende Unternehmen **academicjournals.org**.

Die Situation in den anderen sogenannten Entwicklungsländern dürfte Ähnlichkeiten zu der im subssaharischen Afrika aufweisen, ohne hier dem Vorwurf der *Reductive Repetition* entsprechen zu wollen. (**Andreasson 2005**)

Es ist anzunehmen, dass zukünftig die Zahl der privat organisierten Repositoryn in diesen Ländern zunehmen wird. Die Mehrzahl der Veröffentlichungen darin wird aber weiterhin nicht bzw. nicht ausreichend begutachtet sein.

2 Qualität und Qualitätsfilter

Die Norm ISO 9000:2005 definiert Qualität als „*the degree to which a set of inherent characteristics fulfils requirements*“ (zitiert nach **Vullo et al. 2010, 3**)

Das wesentliche inhärente Merkmal einer wissenschaftlichen Publikation ist die darin enthaltene objektive, gültige und nachvollziehbare Information. Die Qualität dieser Information kann durch Messungen, Überprüfungen, usw. bestätigt oder aber widerlegt werden. Diese Form der Überprüfung ist jedoch aufwändig und in vielen Fällen vom einzelnen Leser auch gar nicht durchführbar. Aus diesem Grund hat sich in der Kultur wissenschaftlicher Publika-

tionen u. a. ein Vertrauensprinzip etabliert, das die Überprüfung des Inhalts durch die Überprüfung der Quelle zu ersetzen versucht.

Aber selbst, wenn der Nachweis von Objektivität, Validität und Reliabilität einer Publikation erbracht werden kann, sind noch nicht alle Aspekte der zuvor zitierten Definition von Qualität erfüllt.

Bezogen auf den Prozess der wissenschaftlichen Kommunikation ergibt sich die Konsequenz, dass der Leser auch seine spezifischen Erwartungen an die Publikation erfüllt sehen muss. Hier spielt die subjektive Relevanz eine Rolle, die der Benutzer einer soeben gelesenen Publikation beimisst.

Diese beide Auffassungen von Qualität sollen im weiteren als **inhaltliche Qualität** bzw. **relevante Qualität** bezeichnet werden. Beide haben innerhalb des Prozesses des wissenschaftlichen Qualitätsmanagements verschiedene Instrumente und Methoden hervorgebracht. Die Aufgabe dieser Qualitätsfilter ist es, jeweils einer der beiden Auffassungen von Qualität gerecht zu werden. Alle nachfolgend beschriebenen Filter tun dies, jedoch können sie nur selten beide Aspekte miteinander verbinden.

Die Begriffe Werkzeug, Instrument und Methode werden auf den folgenden Seiten weitgehend synonym verwendet. Am ehesten schert noch der Begriff der Methode aus dieser Gruppe aus, da hier der dahinter stehende Prozess betont werden soll.

Wir können vier Klassen von Filtern unterscheiden: Begutachtung, Zitationsanalyse, Downloadanalyse und sogenannte Altmetrics. (**Priem et al, 2010**)

Begutachtung in Form von *Peer Review* ist sicher die älteste Form, geplante Veröffentlichungen zu filtern. Es ist für diesen Filter typisch, dass deren Ergebnis in das Wesen der Veröffentlichung eingreift. Ein durchgefallener Text wird nicht veröffentlicht und steht damit der Öffentlichkeit nicht zur Verfügung. Dadurch wird in letzter Konsequenz wissenschaftlicher Erkenntnisgewinn erschwert, z. B. wenn die zu begutachtende Arbeit einen Paradigmenwechsel einleiten könnte. Fehlende Transparenz der klassischen Begutachtung sorgt ebenfalls für Kritik.

Andererseits gibt es keinen anderen Filter, welcher derart aufwändig zu realisieren ist. Führt der Gutachter seine Arbeit korrekt aus, so muss er zunächst die Arbeit lesen. Er muss in den verwendeten Quellen nachschlagen, falls diese noch nicht bekannt sind. Er muss die verwendete Methode überprüfen und schließlich sein Gutachten auch noch in Textform fassen. Da mindestens zwei Gutachter bestellt sein sollten, vervielfacht sich der Aufwand entsprechend. So wurden 2012 für die 116 Zeitschriften der *Nature Publishing Group* nach eigenen Angaben insgesamt 22.454 Gutachter tätig (**Nature 2013**). Die eingehenden Gutachten müssen außerdem noch geprüft und verwaltet werden. Dieses aufwändige Verfahren verspricht zwar effektiv, jedoch nicht effizient zu sein.

Nichtsdestoweniger scheinen jedoch die Vorteile von *Peer Review* zu überwiegen. Ullrich Herb konstatiert dem Verfahren sogar einer Zunahme seiner Verbreitung (**Herb 2012, 21**). Der Hauptvorteil von *Peer Review* ist der damit verbundene Versuch, inhaltliche Qualität zu messen. Auf diesen Vorteil baut im wesentlichen auch *Open Review* auf.

Unter dem Begriff *Open Review* wird eine Reihe von Ansätzen zusammengefasst, die zum Ziel haben, die beschriebenen Nachteile von *Peer Review* durch Transparenz und die Mitarbeit einer möglichst breiten Community aufzuheben. Hierdurch sollen insbesondere soziale Verzerrungen minimiert werden, so wie sie bei der Kritik traditioneller *Peer Review* immer wieder beschrieben werden. Im einfachsten Fall werden die Gutachten veröffentlicht. In der radikalsten Form entspricht *Open Review* dem bekannten Wikipedia-Prinzip, indem jeder die Möglichkeit erhält, die eingereichten Beiträge zu bearbeiten.

Die Ergebnisse sind durchwachsen und hängen insbesondere vom Grad der Mitarbeit durch die Gemeinschaft ab. Diese Mitarbeit kann im Wesentlichen durch zwei Arten von Faktoren erschwert werden. Zum einen gilt es, technische Hürden zu meistern. Zur Mitarbeit ist es in der Regel notwendig sich in einem Forum anzumelden. Ist die Fachdiskussion auf viele Orte verteilt, vervielfacht sich der Aufwand entsprechend. Darüber hinaus muss man sich zur Beurteilung auf eine Webseite zu begeben, die man in der Regel einzig zu diesem Zweck aufsucht. Kommentare z. B. im E-Mail-Client zu verfassen, ist nicht üblich.

Zum anderen spielen auch soziale Faktoren eine Rolle. Während die Mitarbeit an Peer Review noch durch die erhaltene Aufmerksamkeit seitens des Verlages motiviert sein könnte, wird bei der Open Review deutlich, dass diese Art von wissenschaftlicher Arbeit von der Gemeinschaft üblicherweise nicht besonders honoriert wird. Ein weiterer Grund ist die verbreitete Angst, dass das öffentliche Einbringen eigener Ideen in die Beiträge von Kollegen persönliche Nachteile mit sich bringt. Dieser Angst kann dadurch begegnet werden, dass Kommentare und Anmerkungen den gleichen Standards von Veröffentlichungen unterliegen, wie dies auch für die Beiträge selbst gilt. Das *Public Peer Review* von **Copernicus Publications** ist ein Beispiel hierfür. Die Diskussionsforen verfügen dort analog zu den Zeitschriften über eigene ISSN, werden langzeitarchiviert und sind zitierbar. Dies mag ein Grund für den allgemeinen Erfolg der vom Verlag herausgegeben Zeitschriften sein, allen voran *Atmospheric Chemistry and Physics* (**Herb 2012, 26**)

Ein bekanntes Beispiel für ein letztendlich abgebrochenes Open-Review-Projekt ist das *Nature's peer review trial* im Jahre 2006. Grund für den Abbruch war laut Verlag die sehr geringe Beteiligung durch Autoren und Kommentatoren (**Clarke 2006**)

Die **Zitationsanalyse** ist – wie die noch folgenden – ein nachgeordneter Filter. Bereits publizierte Arbeiten werden hinsichtlich der in ihnen verwendeten Referenzen bzw. der auf ihnen verweisenden Zitationen hin untersucht. Leider wird immer wieder die Zahl der erhaltenden Zitationen mit einer hohen inhaltlichen Qualität der Arbeit gleichgesetzt. Im Umkehrschluss wird weniger zitierten Arbeiten eine geringere inhaltliche Qualität unterstellt. Beides stimmt so nicht. Wir können aufgrund einfacher Zählungen nicht auf den Inhalt schließen, sondern nur auf das hervorgerufene Interesse bei den Lesern, also auf die relevante Qualität. Aussagen über die inhaltliche Qualität einer Arbeit sind nur in Kombination mit anderen Methoden möglich. Nicht zitierte Arbeiten werden durch eine prospektive Zitationsanalyse gar nicht erst erfasst. Gerade bei aktuellen Veröffentlichungen ist das in der Regel der Fall. Der Filter versagt hier schlichtweg seinen Dienst.

Verfolgt man die aktuelle Literatur zu Fragen der Qualitätssicherung wissenschaftlicher Publikationen, wird die Kritik an „traditionellen“ Instrumenten der Zitationsanalyse als eine weit verbreitete Gemeinsamkeit deutlich. Insbesondere der *Journal Impact Factor (JIF)* wird massiv in Frage gestellt. Als Beispiel sei hier auf die Arbeit von Ulrich Herb verwiesen, der eine grobe Übersicht über verschiedene Studien gibt, die die Unzulänglichkeiten des *JIF* zum Gegenstand haben. (**Herb 2010, 2**)

Neben der Kritik an der Methode wird auch darauf eingegangen, dass der *JIF* nicht transparent ist (**Arnold et al. 2010**). Fehlende Transparenz jedoch steht dem Konzept von Open Access entgegen.

Downloadanalysen und andere **Usage-Statistiken** berücksichtigen auch jene Arbeiten, die später nicht in Referenzlisten auftauchen. Bereits kurz nach der Veröffentlichung lassen sich Schlüsse auf die Popularität einer Arbeit ziehen. Zudem können Beziehungen zwischen Dokumenten vermutet werden, die in zeitlicher Nähe vom selben Browser aus heruntergeladen wurden. Sabine Henneberger macht in ihrer Dissertation aber auch zu Recht deutlich, wo die Grenzen dieses Filters liegen. Denn einerseits lassen sich nicht alle automatisierten

Downloads aus der Analyse entfernen, andererseits deuten auch Downloadzahlen – ähnlich wie Zitationszahlen – lediglich auf ein Interesse der Leser an der Arbeit, nicht aber auf deren inhaltliche Qualität hin. (Henneberger 2011, 152)

Es gibt noch eine weitere Einschränkung. Wenn wir uns vergegenwärtigen, warum ein bestimmtes Dokument heruntergeladen wird, stoßen wir auf zwei Hauptgründe. Entweder der Nutzer entscheidet sich zum Download, nachdem er lediglich die Metadaten gelesen hat, oder aber er wird durch eine Referenz auf die Arbeit hingewiesen. Der Download selbst geschieht aber in jedem Fall, bevor der Leser die eigentliche Arbeit zu Gesicht bekommt. Unter diesen Umständen kann man schwerlich erwarten, in der Anzahl der Downloads einen funktionierenden Indikator für die inhaltliche Qualität einer Arbeit vorzufinden.

Einen Schritt hin zur quantitativen Analyse inhaltlicher Qualität bildet der **Altmetrics-Ansatz**. Er bezieht bewusst mehrere Indikatoren in die Bewertung ein, darunter auch aktiv durch Benutzer mittels *Social Bookmarking* und *Open Review* erzeugte Daten. Die Idee dahinter ist, dass Benutzer mehr über die inhaltliche Qualität einer Arbeit mitteilen können, als sie dies durch *Usage* und Zitationen bereits tun. Nicht zufällig gibt es einen Zusammenhang zum Wikipedia-Ansatz, einer der Koautoren des **Altmetrics-Manifesto**, Dario Taraborelli, ist zugleich für die Wikimedia Foundation tätig.

Ein viel versprechendes Instrument, welches verschiedene *Altmetrics*-Indikatoren zusammenfasst und veranschaulicht, ist das aus einem bezeichnenderweise *Beyond Impact* genannten Workshop hervorgegangene Projekt **ImpactStory**. Die Seite ging im September 2012 als Beta-Version online. Das Projekt bezieht mit Stand Mai 2013 insgesamt 49 frei verfügbare webometrische Indikatoren aus insgesamt 16 verschiedenen Quellen mit in die Bewertung ein. Alle verwendeten Indikatoren sind quantitativer Natur.

Die Probleme von *Altmetrics* sind vielfältig und lassen sich grob in fünf Typen einteilen:

Die Werkzeuge sind auf möglichst **einheitliche Schnittstellen** zu den verschiedenen Onlinediensten angewiesen. Die Implementierung jeder neuen Schnittstelle ist mit teils erheblichen Aufwand verbunden. Die bisher eingebundenen Schnittstellen müssen ständig überprüft und gegebenenfalls angepasst werden. Die Einbindung aller denkbaren Onlinedienste scheint praktisch unmöglich, wenn man z. B. an die vielen Online-Netzwerke denkt, die mehr auf nationaler oder regionaler Ebene angesiedelt sind. Andererseits lassen sich verlässliche Aussagen zur Qualität einer Arbeit nur herleiten, wenn die Daten aus möglichst vielen unterschiedlichen Quellen bezogen werden.

Die Werkzeuge sind in zweierlei Hinsicht auf **offene Schnittstellen** angewiesen. Zum einen meint offen, dass die Daten der Online-Dienste über API zugänglich gemacht werden; zum anderen, dass dieselben Daten auch ohne Beschränkungen verwendet und weitergegeben werden können.

Die Werkzeuge sind auf die Nutzung einheitlicher **Identifikatoren** seitens der Onlinedienste angewiesen. Dies betrifft insbesondere die eindeutige Unterscheidung von Autoren. Mit der *Open Researcher and Contributor ID (ORCID)* und dem *International Standard Name Identifier (ISNI)* stehen hierzu seit kurzem zwei Möglichkeiten zur Verfügung. Deren umfassende nachträgliche Übernahme in die Repositorien dürfte jedoch nicht einfach sein. Auf ein weiteres Problem mit Indikatoren verweist das Team von *ImpactStory* in seinen FAQ. Eine Arbeit verfügt nur selten über einen einzigen Identifikator. Neben URN und DOI werden auch die URL bzw. *PubMed ID*, *PubMed Central ID*, *Mendeley ID*, usw. verwendet.

Die „Übersetzung“ von **qualitativen Bewertungen** in quantitative Indikatoren gestaltet sich derzeit noch schwierig. Zum Beispiel kann ein Tweet, der eine DOI beinhaltet, zwar als „Erwähnung“ des Beitrages gezählt werden. Jedoch ist es derzeit kaum möglich, aus dem im

übrigen mit vielen nichtstandardisierten Abkürzungen versehenen Tweet eine Bewertung herauszulesen. *ImpactStory* verwendet derzeit – wie bereits erwähnt – ausschließlich quantitative Indikatoren. Die meisten dieser Indikatoren basieren auf *Usage* und der Zahl von Zitationen. Aber es findet sich z. B. auch die Zahl der *Recomendations* durch die ***Faculty of 1000***.

Das meines Erachtens wichtigste Problem ist jedoch die **Mitarbeit der Benutzer**. Sie müssen die auszuwertenden Daten zunächst einmal erzeugen. Nicht jeder Wissenschaftler ist ein Fan von *Twitter* oder *Facebook*. Auch *CiteULike* und *Mendeley* mögen sich in bestimmten Fachdisziplinen noch nicht allzu sehr herumgesprochen haben. Die Nichtteilnahme an Online-Netzwerken wirkt sich in zweierlei Hinsicht aus. Einerseits laufen die eigenen Beiträge Gefahr, nicht ausreichend gewürdigt zu werden. Andererseits ist die weitaus größere Zahl der für die eigene Arbeit hinzugezogenen Literatur von der Erfassung ausgeschlossen. Ansätze wie bei der *Public Peer Review* von *Copernicus Publications* lassen sich nur bedingt übertragen, denn ein Facebook-Like lässt sich nun mal kaum als eigenständige wissenschaftliche Veröffentlichung werten.

Das Dilemma der derzeit ungenügenden Datenlage lässt sich erahnen, wenn man sich die Erläuterungen zu den ermittelten Werten auf der Beispielseite (***sample report***) von *ImpactStory* näher anschaut. So konnte man am 12.5.2013 als Erläuterung zum Tag *Highly saved* eines Artikels (**doi:10.1186/1756-0500-4-487**) nachlesen: „*This item has 3 CiteULike bookmarks. That's better than 97 % of items indexed by Web of Science in 2011, suggesting it's highly saved by scholars.*“ Die Tatsache, dass drei Bookmarks bereits ausreichen, einen Artikel als stark beachtet zu bezeichnen, deutet im besten Fall darauf hin, dass die Blütezeit des *Altmetrics*-Ansatzes in der Zukunft zu suchen ist.

Tabelle 1 fasst die verschiedenen Filter sowie ihre wesentlichen Vor- und Nachteile nochmals zusammen.

Wenn sich eine Tendenz daraus ableiten lässt, dann die, dass die Aufgabe darin besteht, mindestens die Aussagequalität von Peer Review zu erreichen. Das Ganze soll zudem mit geringeren Ressourceneinsatz verbunden und auf die Gesamtzahl von Fächern und Veröffentlichungsformen anwendbar sein.

Der ehemalige Herausgeber von *Nature*, Charles G. Jennings, hat die Hürden aufgezählt, die seines Erachtens jede neue Methode zur Qualitätsbestimmung überwinden muss, um der klassischen Peer Review zumindest ebenbürtig zu sein (**Jennings 2006**). Dazu gehören:

- Reliabilität – die Vorhersage über die Qualität einer Arbeit muss genauer ausfallen, als dies mit derzeitigen Methoden der Fall ist
- Ergebniskomprimierung – die mittels der jeweiligen Methoden getroffenen Aussagen müssen so konzentriert vorliegen, dass leicht ein Überblick gewonnen werden kann.
- Kosteneffizienz – die Aussagen müssen möglichst kostengünstig erzeugt werden können. Dazu gehört auch der Zeitaufwand für die Gutachter.
- Schnelligkeit – die Aussagen müssen kurzfristig verfügbar sein.
- Schutz vor Manipulation – Autoren oder Einrichtungen dürfen nicht in Lage sein, unerlaubten Einfluss auf die Ergebnisse auszuüben.

Bedenkt man, dass Jennings diese Forderungen aus der Perspektive eines sehr erfolgreichen Verlages für kostenpflichtige Zeitschriften formuliert, wird der daraus resultierende Anspruch für Verfechter der Open-Access-Kultur deutlich.

Die vorliegende Arbeit versucht nun, dem an sich erfolgversprechenden *Altmetrics*-Ansatz ein Messinstrument hinzuzufügen, welches noch stärker die qualitativen Aspekte einer wissenschaftlichen Veröffentlichung berücksichtigt.

		Vorteile	Nachteile
Begutachtung	Peer Review	<i>Ermöglicht inhaltliche Qualitätsbestimmung.</i>	<i>Hoher Ressourceneinsatz. Durchgefallene Arbeiten werden nicht veröffentlicht.</i>
	Open Review	<i>Vermindert den negativen Einfluss von sozialen Faktoren im Review-Prozess.</i>	<i>Hoher Ressourceneinsatz. Ist in bestimmten Varianten von der Mitarbeit der Gemeinschaft und von deren gesellschaftlichen Anerkennung abhängig</i>
Zitationsanalyse		<i>Einfache quantitative Bestimmungen sind möglich.</i>	<i>Unzulänglichkeit der Werkzeuge (z. B. Zitationsdatenbanken) Unzulänglichkeit der Methoden (z. B. JIF, h-Index) Keine Analyse von Neuerscheinungen möglich Nur relevante Qualität ist bestimmbar.</i>
Usage-/Downloadanalyse		<i>Einfache quantitative Bestimmungen sind möglich.</i>	<i>Technische Manipulation kann nicht ausgeschlossen werden. Analyse ist begrenzt auf einzelne Angebote (z. B. bestimmte Repositorien). Die Methoden können derzeit noch nicht auf den gesamten Bereich der wissenschaftlichen Veröffentlichungen angewendet werden. Nur relevante Qualität ist bestimmbar.</i>
Altmetrics		<i>Bietet die Möglichkeit, verschiedene Indikatoren miteinander zu vergleichen. Ermöglicht teilweise inhaltliche Qualitätsbestimmung.</i>	<i>Derzeit ist Datenlage noch ungenügend, bedingt durch die geringe Mitarbeit der Gemeinschaft in Online-Netzwerken.</i>

Tabelle 1: Übersicht zu den derzeitigen Filterklassen

3 Ausgangsidee

Die bisherigen Überlegungen weiterführend soll hier ein Instrument vorgeschlagen werden, das zugleich Aussagen zur inhaltlichen als auch zur relevanten Qualität einer wissenschaftlichen Arbeit ermöglicht. Aus Gründen der Vereinfachung trägt das Instrument den Arbeitstitel **OA_SCORE**.

Kern der Methode ist aus Sicht des Lesers eine standardisierte Kurzbewertung, die dieser unmittelbar nach der Lektüre der Veröffentlichung abgeben kann. Es wird vermutet, dass das zusammengefasste Urteil einer möglichst großen Anzahl von Lesern gültige Aussagen zur inhaltlichen Qualität einer Arbeit zulässt. Diese Herangehensweise schließt auch jenen Kreis von Lesern ein, die die Arbeit später nicht mehr explizit erwähnen werden. Das Instrument versteht sich als Ergänzung zum *Altmetrics*-Ansatz.

Die Auswertung soll durch einen zentralen Service-Provider erfolgen. Die Ergebnisse werden den *Data Provider* zur Verfügung gestellt, damit sie zusammen mit den Metadaten angezeigt werden können. Gleichzeitig können sie anderen Service-Providern zum Ranking von Ergebnislisten dienen.

Ein einfaches Entity-Relationship-Modell von OA_SCORE ist in **Abbildung 1** dargestellt.

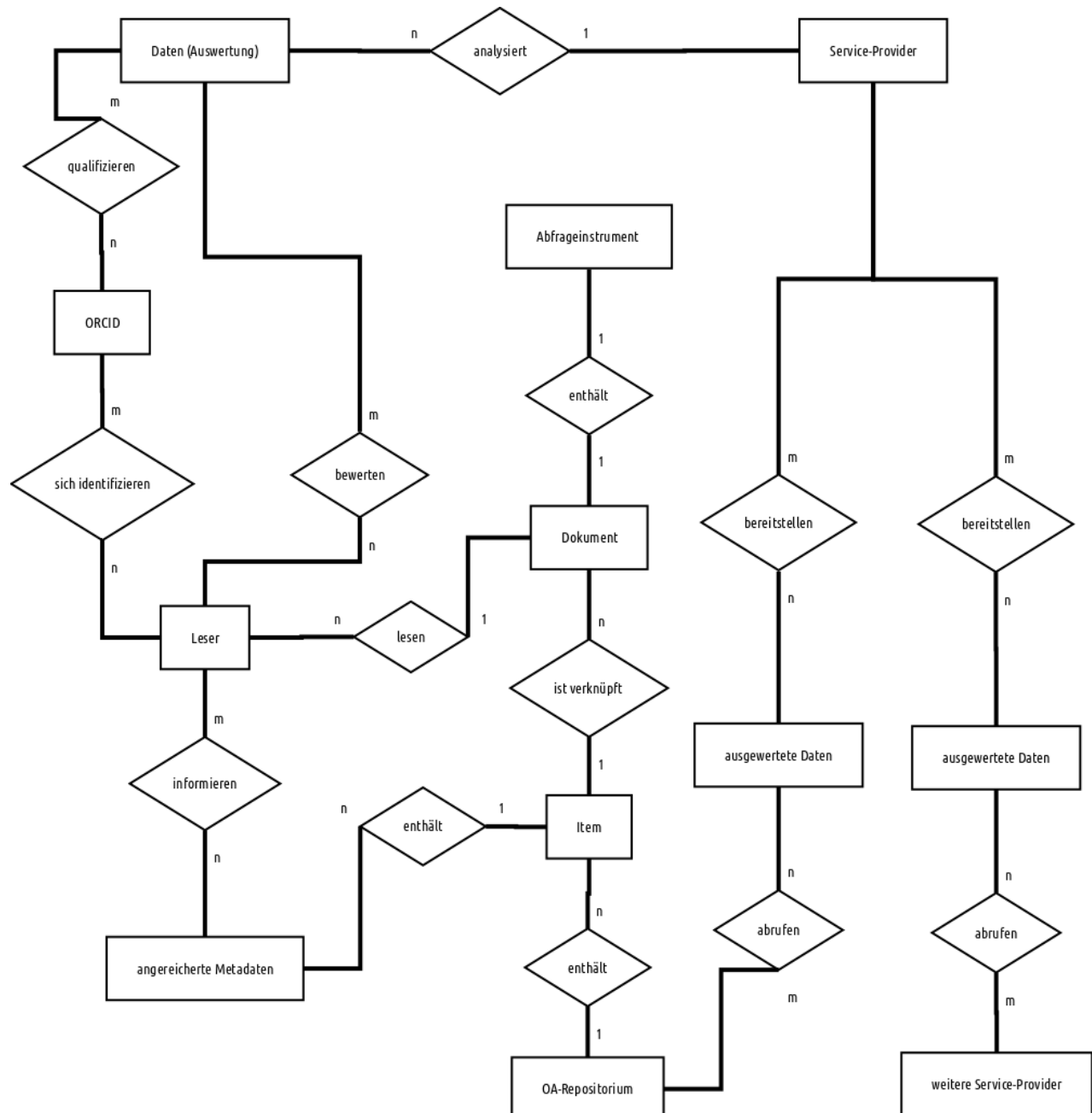


Abbildung 1: ER-Modell von OA_SCORE

Das Prinzip, das Leser Online-Beiträge in standardisierter Form kurz bewerten, ist im *World Wide Web* weit verbreitet. Häufig kann hierzu eine bestimmte Anzahl von Sternen oder Punkten vergeben werden. Diese werden dann zusammengefasst und sollen den anderen Lesern einen Eindruck vom Beitrag geben bzw. in das Ranking von Suchergebnissen mit

einfließen. Das Verfahren ist problematisch. Zum einen existieren bei den Bewertern unterschiedliche Auffassungen zum Wesen des häufig eindimensionalen Bewertungskriteriums. „Gut“ und „Gut“ haben in den Augen verschiedener Betrachter mitunter verschiedene Bedeutungen.

Zum anderen müssen die potentiellen Bewerber bereit sein, die Bewertung durchzuführen. Diese Bereitschaft hängt u. a. von zwei Aspekten ab:

- Das **Verhältnis zwischen Teilnahmeaufwand und erwartetem Nutzen**. Geringer Teilnahmeaufwand birgt die Gefahr der unscharfen Aussage. Beispielsweise gibt eine einzelne eindimensionale Skala keine Aussage darüber, warum ein Objekt gut, weniger gut oder schlecht ist. Werden hingegen komplexere Abfragen durchgeführt, sinkt tendenziell die Bereitschaft zur Mitarbeit auf Seiten der Zielgruppe, wenn sie gleichzeitig keine sonstige Art von Anerkennung erlangt.
- Die **Länge des Weges zwischen Objekterfahrung und Bewertung**. Hiermit ist die Barriere gemeint, die sich ergibt, wenn der Nutzer eine Bewertung abgeben möchte. Im günstigsten Fall ist dieser Weg kaum messbar, z. B. eine einfache anonyme Kommentarfunktion am Ende eines Webseitentextes. Der Weg verlängert sich, wenn sich der Nutzer zum Zwecke des Kommentierens erst registrieren muss oder wenn der Ort der Objekterfahrung nicht identisch mit dem Ort der Bewertung ist. Dies ist z. B. bei der Bewertung eines im separaten Betrachter geöffneten PDF-Dokuments über eine Webseite im Browser der Fall.

Aus den genannten Punkten ergeben sich folgende Bedingungen für den Einsatz des zu konzipierenden Instruments:

- Die Bereitschaft zur Mitarbeit kann nur durch die größtmögliche Einfachheit der Methode erreicht werden. Der zusätzliche Aufwand für den Leser muss minimal sein. Der Aufwand kann im Interesse einer genaueren Aussage höher sein, wenn dem Leser darüber hinaus zusätzlicher Nutzen angeboten wird.
- Die Bewertung muss trotzdem mehrere Objekteigenschaften abfragen, um dadurch deren Aussagekraft zu erhöhen.
- Die Bewertung muss im Dokument erfolgen, um die Wege kurz zu halten.

Weitere Bedingungen sind:

- Die Methode muss auf einen großen Teil der in OA-Repositorien veröffentlichten Dokumente anwendbar sein.
- Die Methode muss eine maschinelle Auswertung der gesammelten Bewertungen ermöglichen.
- Die Methode muss gültige und nachvollziehbare Aussagen ermöglichen.
- Die Methode muss fächerübergreifend anwendbar sein.
- Die Methode muss interkulturell anwendbar sein
- Die Methode muss quelloffen dokumentiert sein.
- Die erzeugten Daten müssen mit einer Lizenz versehen sein, die die Weiterverwendung durch Dritte gewährleistet.
- Die Methode muss weitgehend resistent gegen Manipulationsversuche ausgelegt sein.

Die letzte Bedingung des Manipulationsschutzes ist auch der Grund dafür, dass im Vorschlag ein zentraler Service-Provider zwischengeschaltet ist. Über ihn soll mittels Identitäts-

abfrage eine Mehrfachbewertung durch die gleiche Person ausgeschlossen werden. Es ist im Sinne der Forderung nach einfacher Anwendbarkeit undenkbar, dass jedes Repositorium die Identitäten seiner Leser verwaltet.

Es wird angenommen, dass die meisten zu bewertenden Dokumente technisch die Voraussetzungen für die Einbettung eines Auswertungstools bieten. Ein solches Dateiformat stellt zum Beispiel das *Portable Data Format (PDF)* dar. Es besitzt mehrere vorteilhafte Eigenschaften:

- Dem Dokument können im Moment der Auslieferung zusätzliche Elemente automatisch beigelegt werden.
- Das Dokument erlaubt die Einbettung von Links, welche zudem über Grafiken angesprochen werden können.
- Es handelt sich um einen offenen Standard, der durch ISO 32000-1 dokumentiert ist. Zudem bietet die Entwicklerfirma Adobe® eine kostenlose Version der Dokumentation an (**Adobe 2008**)

Die Bewertung selbst soll mittels eines *HTTP-GET-Request* an den Service-Provider realisiert werden, welcher die abgefragten Daten im *Query String* der URL transportiert.

Anstatt in ein PDF-Dokument ließe sich das Instrument auch in andere Dateiformate integrieren, z. B. *text/html* oder *text/xml*.

Im Rahmen dieser Arbeit liegt der Schwerpunkt jedoch auf dem *Portable Data Format (PDF)*. Es stellt eine wesentliche technische Grundlage für OA SCORE dar. Das Format scheint eine besondere Rolle in der Wissenschaftskommunikation zu spielen, wenngleich verschiedene Initiativen bereits alternative Kommunikationsformate für die Zukunft in die Diskussion einbringen. Ein Beispiel hierfür ist die von der virtuellen Fach-Community *Future of Research Communications and e-Scholarship (Force11)* organisierte Workshopreihe **Beyond the PDF**. Diese Initiative ist sicherlich zu begrüßen, versucht sie doch letztendlich die Gesamtheit der Wissenschaftskommunikation an die aktuellen technischen Möglichkeiten anzupassen. Dass jedoch PDF dabei in gewisser Weise als das zu überwindende Hindernis postuliert wird, scheint meines Erachtens zu weit gefasst zu sein. Die technischen Möglichkeiten, die PDF heute bereits bietet, werden bisher kaum angewendet. Zudem wird die Entwicklung des Formats vorangetrieben und steht seit 2012 mit PDF/A-2 ein erweiterter Standard zur Archivierung bereit (**Drümmer 2011, 497**)

Ein positives Beispiel, wie Interaktivität in der Wissenschaftskommunikation mittels PDF eingesetzt werden kann, findet sich in den Artikeldiskussionen der Zeitschriften von *Copernikus Publikations*. Der Leser findet am Rand des PDF-Dokuments Links, die ihm u. a. bei der Durchsicht des Artikels helfen.

Auch innerhalb des von David Shotton (**Shotton 2012**) vorgeschlagenen 5-Sterne-Ansatzes zur Evaluation von Online-Artikeln findet sich durchaus ein Platz für PDF. Die im Unterpunkt *enriched content* angeführte fünfstufige Skala kann durch PDF bis zur dritten Stufe *Semantic enrichment of the text* bedient werden. PDF/A-2 erlaubt dies. Mit dem geplanten Standard PDF/A-3 wird dann sicher auch der vierte Unterpunkt *'Lively' content* und eventuell auch der fünfte Unterpunkt *Data fusions* erfüllbar sein (**Drümmer 2011, 497**).

PDF wird so auch in Zukunft eine wichtige Rolle in der Wissenschaftskommunikation spielen können.

4 Forschungsfragen

Im Rahmen dieser Arbeit können nicht alle zuvor genannten Bedingungen mit wissenschaftlichen Methoden überprüft werden.

Ich beschränke mich deshalb auf die Untersuchung der folgenden Aspekte.

- **Technische Anwendbarkeit**
Sind die technischen Voraussetzungen gegeben, damit der Filter angewendet werden kann? Hier sollen nicht die technischen Spezifikationen hinsichtlich einer Implementierung von OA_SCORE untersucht werden. Vielmehr geht es um die Frage, ob es in den OA-Repositorien eine ausreichende Anzahl von Dokumenten gibt, die sich mittels dieser Methode technisch auswerten lassen.
- **Bedarfsermittlung**
Wie viele Dokumente werden in OA-Repositorien angeboten, deren inhaltliche Qualität nicht schon über andere gängige Filter ausreichend gut ermittelt werden konnte? Die Frage zielt auf die Ermittlung des Anteils von nicht-begutachteten Veröffentlichungen in Open-Access-Repositorien ab.
- **Plausibilität der gewonnenen Aussagen**
Können mit dem neuen Instrument plausible Aussagen getroffen werden? Damit einher geht die Frage, ob das Instrument für die Anwender ausreichend verständlich ist.

Darüber hinaus werden Fragen zum Design zu klären sein. Dies betrifft u. a.:

- Die Form der Implementierung
- Die Form der Abfrage
- Die Art und Weise, wie die gewonnenen Daten für andere Dienste aufbereitet werden können.
- Der Schutz vor Manipulationen

5 Methoden

5.1 Technische Anwendbarkeit

Um die Frage der technischen Anwendbarkeit zu klären, muss die Anzahl der Open-Access-Dokumente ermittelt werden, deren Dateityp eine Implementierung von OA_SCORE zulässt. Außerdem muss geklärt werden, inwiefern die Repositoriums-Software modular aufgebaut ist und eine serverseitige Implementierung des Instruments erlaubt.

Die **Nullhypothese** besagt, dass die aktuelle durch OA-Repositorien vermittelte wissenschaftliche Kommunikation mehrheitlich nicht mit Hilfe von PDF-Dateien erfolgt.

Der in dieser Arbeit verwendete Begriff des **OA-Repositoriums** bezieht sich ausschließlich auf die Gruppe der in *ROAR* bzw. *OpenDOAR* gelisteten Repositorien. Es ist bekannt, dass die Gesamtzahl aller Repositorien mit frei zugänglichen Inhalten bedeutend größer ist.

Die eingesetzte Methode zur Widerlegung der Nullhypothese ist eine Zählung der mittels *Metadata Harvesting* gewonnene Formatangaben zu den einzelnen Items und Dokumenten.

Hierfür wird das von der *Open Archives Initiative* veröffentlichte *Protocol of Metadata Harvesting in der Version 2.0* (OAI-PMH) verwendet (**Lagoze et al., 2002**). Die Liste der untersuchten Repositorien stellt die Summe der Einträge des *Open Directory of Open Access Repositories* (**OpenDOAR**) und des *Registry of Open Access Repositories* (**ROAR**) dar.

Im weiteren Verlauf der Arbeit übernehme ich die von der *Open Archives Initiative* vorgeschlagene Unterscheidung zwischen *Ressource*, *Item* und *Document*.

Eine **Ressource** ist hier ein Informationsobjekt, welches physisch oder digital vorliegen kann. Dies können neben digitalen Daten auch Bilder, Bücher, Museumsobjekte, usw. sein. Bei den im Rahmen von OAI-PMH verwalteten Daten handelt es sich um Metadaten in Bezug auf eine oder mehrere Ressourcen.

Ein **Item** stellt eine abgeschlossene Einheit von Metadaten innerhalb eines Repositoriums dar und beschreibt eine Ressource bzw. eine Gruppe von zusammenhängenden Ressourcen.

Ein Item kann kein, ein oder mehrere elektronische **Dokumente** enthalten. Hierbei handelt es sich in der Regel um Dateien, welche in verschiedenen Formaten bzw. Dateitypen vorliegen können. Diese Dokument-Definition ist also lediglich eine Teilmenge der von Jacco van Ossenbruggen vorgeschlagenen Definition, wonach ein Dokument als „*a self-contained unit of information, intended to be communicated to one or more human interpreters.*“ definiert ist (von Ossenbruggen 2001, 10).

Zudem kennt OAI-PMH noch den Begriff des **Record**. Dabei handelt es sich um die Repräsentation eines Items bzw. Teilen davon in einem bestimmten **metadataFormat** in Form eines XML-kodierten Streams.

Die Erhebung hat den Anspruch, die über OAI-PMH zugänglichen Dokumente vollständig einzubeziehen.

Hintergrund ist die Annahme, dass alle in *OpenDOAR* bzw. *ROAR* nachgewiesenen Repositorien mit Hilfe von OAI-PMH ausgewertet werden können, da die entsprechende Schnittstelle in der jeweiligen Repositoriums-Software implementiert ist. Die abgerufenen XML-Streams werden Record für Record darauf getestet, ob Formatangaben enthalten sind, die sich als MIME-Typen identifizieren lassen. Jede zusätzlich ermittelte Formatangabe wird in eine Liste eingetragen. Die Einträge werden durch Angaben zur Häufigkeit ergänzt.

Gezählt werden sollen nur Angaben, die sich auf einen der offiziellen von der **IANA** registrierten MIME-Typen beziehen.

Es wurde versucht, sich bei der Abfrage der Records auf das Metadatenformat *Metadata Encoding and Transmission Standard* (**METS**) zu konzentrieren. Es zählt zu den Austauschformaten, die einen einheitlicheren Datenrücklauf erwarten lassen (Van de Sompel et al. 2004). Bei Repositorien, die Records nicht in *METS* ausliefern, wurde auf das durch OAI-PMH vorgeschriebene Mindestformat *OAI_DC* zurückgegriffen.

Die Daten werden mit Hilfe eines **BASH-Skriptes** gewonnen. Das Skript läuft über einen in Deutschland ansässigen virtuellen Server unter Ubuntu Server 10.04. Für XML-Abfragen wurde **xmlstarlet** verwendet. Da der Server hauptsächlich für das Bearbeiten von HTML-Anfragen benutzt wird, nur wenig Festplattenspeicher besitzt und nur über 2 GB RAM verfügt, konnten übliche Harvesting-Tools nicht eingesetzt werden. Gerade bei umfangreichen Repositorien ist es nicht möglich, zunächst den gesamten XML-Stream zu laden und diesen erst im Anschluss zu untersuchen. Stattdessen werden die einzeln gelieferten Teil-Streams direkt ausgewertet und anschließend verworfen. Die zusammengefassten Daten werden in einer Ergebnis-Datei abgelegt.

Die Konfiguration des Harvesting-Prozesses bringt es mit sich, dass neben der Zahl der in den verschiedenen Formaten vorliegenden Dokumenten auch die Gesamtzahl der Records sowie die Zahl der Records, welche Dokumente enthalten, gezählt werden. Letztere sollen im weiteren Verlauf der Arbeit als **Dokumenten-Records** bezeichnet werden.

Folgende Probleme sind zu erwarten:

Eine **Uneinheitlichkeit der abgerufenen Daten** wird sich nicht vermeiden lassen. Das XML-Schema des OAI-PMH-Protokolls beschreibt lediglich einen Container, der wiederum mit verschiedenen Metadatenformaten gefüllt werden kann.

Von den beiden verwendeten Metadatenformaten bietet *METS* die besseren Chancen, MIME-Typen auszulesen. Innerhalb des *file*-Elements ist die Angabe eines MIME-Typs möglich. Der Standard verweist hier ausdrücklich auf die von der **IANA** aufgestellte Liste gültiger MIME-Typen (**Digital Library Federation 2007, 30**). Daraus folgt, dass mindestens ein digitales Dokument vorliegt, wenn mindestens ein *file*-Element mit MIME-Typ vorliegt. Dies bedeutet allerdings nicht automatisch, dass dies auch die Datei mit dem eigentlichen *Content* ist.

Schwieriger gestaltet sich die Situation beim *OAI_DC*-Format. Die Spezifikation gibt für das *DC:Format*-Element verschiedene Vorschläge an, da hierdurch sowohl physische als auch digitale Ressourcen beschrieben werden können. Für den Fall, dass es sich hier um eine digitale Ressource handelt, empfiehlt die Spezifikation von *Dublin Core* die Verwendung der von der IANA aufgestellten MIME-Typen. Sie schreibt sie aber nicht vor (**Dublin Core 2012**).

Die **Implementierung des Metadatenexports** durch die jeweilige Repositorium-Software ist ebenfalls uneinheitlich gestaltet. Die beiden am häufigsten verwendeten Anwendungen **DSpace** und **EPrints** beinhalten beide von Hause aus die Möglichkeit des Exports im *METS*-Format. Andere Anwendungen sehen dies nicht vor, darunter z. B. auch die spezielle Applikation auf der das *ArXiv*-Repositorium läuft.

Die einzelnen Software-Lösungen gehen auf unterschiedliche Weise mit dem zum Item gehörenden Hilfsdateien um. Gemeint sind hier *Thumbnails*, Index-Dateien, usw. *DSpace* scheint zum Beispiel standardmäßig beim Vorliegen von PDF-Dokumenten auch eine Datei mit dem MIME-Typ *text/plain* zu generieren. Beim **EDOC-Server der HU** wird der Inhalt von PDF-Dokumenten zugleich im HTML-Format angeboten. Zusätzlich wird eine HTML-Datei angeboten, welche lediglich die Metadaten enthält.

Bei nicht wenigen Repositorien finden sich überhaupt keine Formatangaben, obwohl auf der Webseite beim entsprechenden Item offensichtlich Dateien zum Download angeboten werden. In diesem Fall wird die Untersuchung nach 500 Items abgebrochen, falls bis dahin kein Dokument mit Formatangabe gefunden wurde.

Ein weiteres Problem besteht in der inkonsistenten **Dateneingabe durch die Anwender**. Offenbar überlassen einige Applikationen dem Anwender die manuelle Eingabe des Formats. Dadurch entsteht die Gefahr, dass die Angaben nicht mehr eindeutig einem der offiziellen MIME-Typen zugeordnet werden können.

Schließlich muss davon ausgegangen werden, dass die **Erreichbarkeit der OAI-PMH-Schnittstelle** nicht in jedem Fall gegeben ist. So fällt beim Betrachten der Datenbankauszüge von *OpenDOAR* und *ROAR* auf, dass bei vielen Repositorien keine Adresse der Schnittstelle angegeben ist.

Im Zusammenhang mit der Erreichbarkeit der Schnittstelle kann es auch passieren, dass die Schnittstelle das sogenannte *ResumptionToken* nicht erkennt und daraufhin keine Daten mehr ausliefert. Über das *ResumptionToken* können Data-Provider die durch Abfragen entstehende Serverbelastung ausgleichen, indem sie die Zahl der jeweils zurückgegeben Items begrenzen (**Lagoze et al. 2002**).

5.2 Bedarfsermittlung

Auch die Frage, ob durch den Anteil der nicht-begutachteten Items ein Bedarf an OA_SCORE entstehen könnte, soll mittels Datamining geklärt werden.

Von Paul Vierkant gibt es bereits einen aktuellen Überblick über den Anteil der verschiedenen OA-Veröffentlichungsarten (**Vierkant 2012, 59**). Dessen Ergebnis soll mit den hier ermittelten Werten verglichen werden.

Die **Nullhypothesen** besagen in diesem Zusammenhang,

- dass der Anteil nichtbegutachteter Items in OA-Repositorien nicht mit ausreichender Genauigkeit und Vollständigkeit zu ermitteln ist und
- dass es keine signifikante Anzahl an nicht-begutachteten Items in den untersuchten OA-Repositorien gibt.

Die Zählung erfolgt zugleich mit der Zählung der MIME-Typen, allerdings ist es hier weitaus schwieriger, belastbare Aussagen zu erhalten. Während Dateien noch relativ sicher über ihren, einem kontrolliertem Vokabular entstammenden MIME-Typ identifizierbar sind, gilt dies nicht mehr hinsichtlich des Begutachtungsstatus. Begutachtung ist ein Prozess, dessen Durchführung in OA-Repositorien nicht standardisiert erfasst ist.

Es werden nur Items untersucht, bei welchem zuvor bereits das Vorhandensein von MIME-Typen festgestellt wurde. Die Zählung selbst bezieht sich jedoch auf die Dokumente.

Im Folgenden wird davon ausgegangen, dass drei Arten von Veröffentlichungen grundsätzlich begutachtet werden: Zeitschriftenartikel, Beiträge in Sammelbänden und Qualifikationsarbeiten. Es muss also in den einzelnen Datensätzen nach Hinweisen für eine dieser Veröffentlichungsarten gesucht werden. Im Skript wurden deshalb folgende Abfragen implementiert, wobei Artikel und Beiträge für Sammelbände zusammen erfasst werden sollen:

	Abfrage über METS	Abfrage über OAI_DC
Zeitschriften- artikel / Beiträge in Sammelbänden	Inhalt von <setSpec> = *journal* *book* *peer* type-Attribut in <identifier> = isbn issn	Inhalt von <DC:type> = *periodical* *journal* *proceeding* *issn* *isbn* *sage* *springer* *elsevier* *series* 978* *revista* *revue* *zeit- schrift* *acta* *peer* Inhalt von <DC:type>= Pattern einer ISSN (0000-0000 bzw. 0000-000x)
Qualifizierungs- arbeiten	Inhalt von <setSpec> = *tesis* *thesis* *gradua* *dissertatio* Inhalt von <roleTerm> = *advisor* *gutachter* *betreuer* Inhalt von <genre> = *thesis*	Inhalt von <DC:type> = *tesis* *thesis* *mas- ter* *dissertation* *bachelor*
Nicht begutachtet	Inhalt von <setSpec>= *unpublis* *pre- print*	

Tabelle 2: Übersicht über die Suchpattern zur Ermittlung von begutachteten Veröffentlichungen

Aus der Übersicht in **Tabelle 2** wird bereits ersichtlich, dass eine vollständige Erfassung des Anteils begutachteter Items nicht möglich ist. Neben der nicht einheitlichen Kennzeichnung durch ein bestimmtes Element, werden tendenziell auch Hinweise in weniger häufigen Sprachen nicht erfasst.

Ein weiteres Problem ist das Alter der Veröffentlichung. Die Zählung erfolgt über der Gesamtmenge von Items in einem Repositorium, Spezifische Sets – wie von OAI-PMH vorgesehen – wurden nicht berücksichtigt. Dadurch ist es möglich, dass in ein und demselben Repositorium sowohl aktuelle Beiträge als auch historische Dokumente erfasst werden. Ein sol-

ches Beispiel ist das Repositorium der Universität Tartu/Estland, das sowohl aktuelle Beiträge als auch digitalisierte alte Drucke anbietet.

Bei der Untersuchung der Ergebnisse wird wegen der beschriebenen Einschränkungen darauf zu achten sein, dass nur Repositorien mit plausiblen Zählungen darin einfließen.

Als plausibel können die ermittelten Werte zu einem Repositorium betrachtet werden, wenn das Verhältnis der begutachteten Dokumente zur Gesamtzahl der Dokumente signifikant, also größer als 5 % ist. Zugleich darf das Repositorium laut Angabe in *OpenDOAR* nur Beiträge in den Sprachen Englisch, Französisch, Deutsch, Spanisch oder Italienisch anbieten.

In jedem Fall muss nach Abschluss des *Harvesting* nachträglich eine Ergebnisstichprobe manuell gesichtet werden. Dadurch soll sichergestellt werden, dass kein systematischer Fehler vorliegt.

Andere Veröffentlichungsarten, bei denen eventuell Begutachtungen stattfanden, werden nicht erfasst und zählen somit als nicht-begutachtet. Dies betrifft z. B. Forschungsberichte, Konferenzbeiträge und Posterpräsentationen. Der daraus resultierende Fehler muss abgeschätzt werden.

Von den ermittelten Werten kann nicht auf die Gesamtmenge der OA-Veröffentlichungen geschlossen werden. Die Untersuchung kann lediglich versuchen zu belegen, dass es eine signifikante Zahl von nicht-begutachteten Dokumenten in den betrachteten Repositorien gibt und sich dementsprechend der Einsatz von OA_SCORE anbietet. Zugleich soll jedoch auch deutlich werden, dass die Anwendung dieses Instruments nicht in jedem Fall sinnvoll ist.

5.3 *Plausibilität der gewonnenen Aussagen*

Hier geht es darum, das Bewertungsdesign zu entwickeln, welches in OA_SCORE zur Anwendung kommen soll. Es soll auf drei Fragen hin getestet werden:

- Sind die zu bewertenden Kriterien verständlich?
- Ist die Form der Abfrage verständlich?
- Können plausible Aussagen getroffen werden?

Entsprechend lauten die **Nullhypothesen**:

- Es gibt keine universell anwendbaren Kriterien zur Kurzbewertung einer wissenschaftlichen Arbeit.
- Es gibt keine Möglichkeit, plausible Aussagen durch das Instrument OA_SCORE zu erhalten.

5.3.1 **Bewertungsdesign**

Das Bewertungsdesign soll anhand einer standardisierten mündlichen Befragung getestet werden. Die Zielgruppe setzt sich aus Personen mit Abitur und unterschiedlichen kulturellen und muttersprachlichen Hintergrund zusammen.

Eine Besonderheit des Designs besteht darin, dass OA_SCORE aus technischen Gründen nur eine einzige Antwort zulässt. Es kommt folglich darauf an, soviel wie möglich bewertende Informationen auf diese Art zu übermitteln. Idealerweise sollte es sich um eine multivariate Information handeln, d. h. der Befragte übermittelt mit einer einzigen Aktion, nämlich dem Klick auf einem Link, mehrere Aussagen. Die einzusetzende Skala muss demzufolge mehrdimensional gestaltet sein. Der Begriff der Mehrdimensionalität bezieht sich hierbei auf die Abfrage mehrerer Variablen durch nur eine Skala.

So weit ersichtlich, finden sich in der Literatur keine Studien, die sich auf Tests mit mehrdimensionalen Eingabeskalen beziehen. Üblicherweise werden bei Umfragen mehrere Variablen nacheinander und nicht gleichzeitig abgefragt.

Eine **zweidimensionale Bewertung** lässt sich grundsätzlich über ein in einem kartesisches Koordinatensystem eingebettetes Diagramm realisieren. Dabei klickt der Befragte auf eine x,y -Koordinate. Der Wert wird anschließend in eine x - und in eine y -Komponente aufgeteilt. Es ist zu vermuten, dass diese Form der Abfrage leicht verständlich ist. Für die Untersuchung erhält jede Dimension eine gerade Anzahl von Ausprägungen. Dies soll verhindern, dass unschlüssige Antworten sich um eine mittlere Ausprägung häufen.

Jedes Kriterium für sich stellt eine Verhältnisskala dar. Die Ausprägung 1 bedeutet, dass das Kriterium nicht vorhanden ist. Da Verhältnisskalen linear transformiert werden können, lässt sich durch Subtraktion von 1 ein Raum $(0,1,2,3)$ für die Merkmalsausprägungen festlegen. **Abbildung 2** zeigt ein Beispiel für eine zweidimensionale Skala.

Abbildung 2:

Beispiel einer zweidimensionalen Skala (2D-Skala)

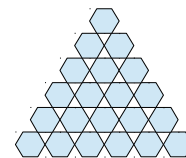
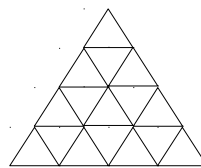
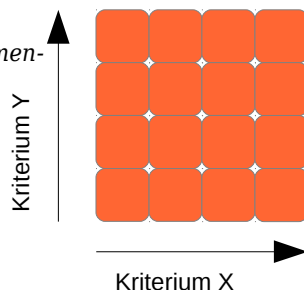


Abbildung 3: Entwicklungsschritte zu einer 3D-Skala

Für eine **dreidimensionale Bewertung** kann das Diagramm theoretisch in ein dreidimensionales kartesisches Koordinatensystem eingebettet werden. Da jedoch die Darstellung auf eine zweidimensionale Fläche beschränkt bleibt, können nicht alle Ausprägungen eindeutig dargestellt werden. Um das Problem zu umgehen, könnten die Winkel zwischen den Achsen verändert und so zusätzlich eine Achse geschaffen werden. Das Ergebnis würde einem gleichseitigem Dreieck ähneln.

Ein Punkt innerhalb dieser Struktur lässt sich durch dessen jeweils kürzesten Abstand zu den drei Seitenlinien definieren. Damit wären prinzipiell dreidimensionale Aussagen möglich. Die beiden in **Abbildung 3** vorgeschlagenen Skalen sind jedoch problematisch. Bei der linken Skala bewirkt die geometrische Form der Dreiecke eine unterschiedliche Wahrnehmung des Abstandes beim Betrachter. Das Orthozentrum eines gleichseitigen Dreiecks stimmt nicht mit dem Mittelpunkt der – vom Betrachter aus gesehen senkrechten – Höhe zusammen. Dadurch kann der Mittelpunkt eines Dreiecks von vielen Menschen schlecht abgeschätzt werden. Um das Problem zu umgehen, könnten regelmäßige Sechsecke als Elemente der Skala verwendet werden. Das Problem hierbei sind die dadurch entstehenden dreieckigen Lücken, die einen nicht bewertbaren Raum innerhalb des Diagramms darstellen. Zudem lässt sich auch hier der Mittelpunkt des gesamten Diagramms schlecht abschätzen. Viele Menschen tendieren wahrscheinlich dazu, ihn näher bei der unteren Grundlinie zu sehen. Dadurch sind Verzerrungen zugunsten des dort markierten Kriteriums zu erwarten.

Es wird deshalb eine andere Form einer 3D-Skala vorgeschlagen, die sich durch einen klar markierten Mittelpunkt auszeichnet (**Abbildung 4**).

Eine Besonderheit der bisher betrachteten dreidimensionalen Skalen besteht darin, dass damit keine Absolutwerte sondern die Beziehung zwischen den Kriterien gemessen werden. Bei **Abbildung 4** entscheidet der Befragte sich zunächst für das Kriterium, welches seiner Meinung nach am meisten zutrifft (grün, rot oder blau). Anschließend wählt er innerhalb

der Felder „seiner“ Farbe dasjenige aus, welches die Abstände zu den beiden anderen Kriterien am besten widerspiegelt. Dem Prinzip nach handelt es sich um eine Ordinalskala. Den drei Kriterien werden Ränge zugewiesen. Dabei gibt es immer einen ersten Rang, während die beiden nachgeordneten Kriterien u.U. auch gleichrangig sein können.

Die Abstände zwischen zwei Kriterien können in der dargestellten Skala die Ausprägungen 1,2 bzw. 3 haben. Es können maximal zwei von drei möglichen Beziehungen erfasst werden. So können beispielsweise bei der Wahl eines grünen Feldes die Abstände grün-rot und grün-blau, nicht jedoch der Abstand rot-blau festgestellt werden.

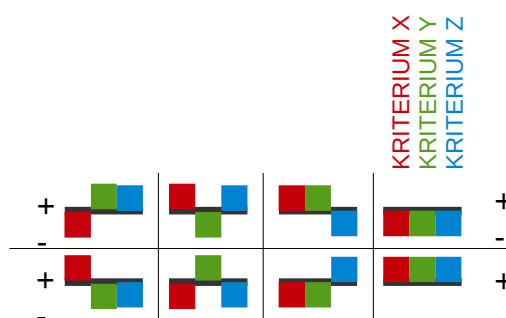
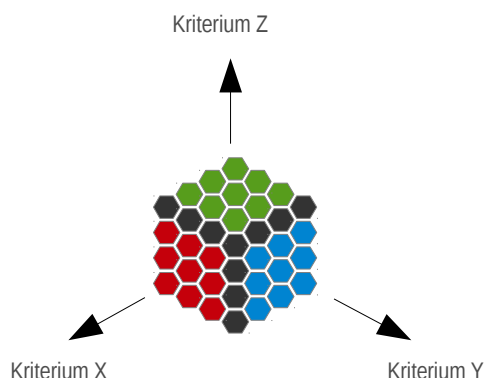


Abbildung 4: Beispiel einer dreidimensionalen Skala (3D-Skala)

Abbildung 5: Beispiel einer 1-von-8-Skala

Der Begriff des Abstandes zwischen zwei unterschiedlichen Kriterien bedarf noch einer kurzen Erläuterung. Gemeint ist hier der gefühlte Unterschied zwischen dem Grad der Ausprägung der einzelnen Kriterien. Dadurch sollen Aussagen möglich werden, wonach z. B. Kriterium 1 ganz, jedoch Kriterium 2 nur teilweise zutrifft. „Abstand“ bedeutet demnach der Unterschied zwischen „trifft ganz zu“ und „trifft teilweise zu“. Ein objektiver Abstand zwischen zwei gänzlich unterschiedlichen Kriterien lässt sich nicht ermitteln.

Es gibt noch eine weitere Möglichkeit, drei Kriterien mit einer Antwort abzufragen. **Abbildung 5** zeigt ein Beispiel. Dem Befragten werden hier alle möglichen Antwortkombinationen aufgezeigt und er wählt eine davon aus. Zutreffende Kriterien werden durch ein Feld oberhalb einer Linie gekennzeichnet. Bei nichtzutreffenden Kriterien befindet sich das Feld unterhalb der Linie. In diesem Fall wäre aber nur eine binäre bzw. dichotome Ausprägung pro Variable praktisch umsetzbar. Bei drei Variablen mit jeweils binärer Ausprägung gibt es somit acht mögliche Kombinationen. Stehen hingegen je Variable vier Werte zur Verfügung, erhöht sich die Zahl der möglichen Kombinationen auf $4^3 = 64$. Auch die Variante mit vier Variablen und binärer Ausprägung dürfte sich als nicht praktikabel erweisen. Es müssten $2^4 = 16$ Kombinationen dargestellt werden.

Variablen mit drei Ausprägungen sind hingegen problematisch. Die mittlere Ausprägung könnte hier nicht nur als Mittelwert, sondern auch als Entscheidungsunwilligkeit interpretiert werden.

Dem Wesen nach handelt es sich bei dieser vorgeschlagenen Skala um eine Kombination von dichotomen Skalen, die über eine Entscheidungsmatrix operationalisiert werden können.

Für die Untersuchung wird eine Skala mit drei Variablen und dichotomer Ausprägung benutzt. Im weiteren Verlauf soll sie vereinfachend als **1-von-8-Skala** bezeichnet werden.

Aufgrund der besonderen Bewertungssituation wurde zunächst die Form der Eingabeskala diskutiert. Selbstverständlich ist es trotzdem notwendig, hier die Frage zu besprechen, welche Indikatoren überhaupt gemessen werden sollen.

Die per Kurzbewertung zugebilligte inhaltliche Qualität einer wissenschaftlichen Arbeit ist eine latente Variable. Sie lässt sich nicht direkt messen. Hinzu kommt, dass sie teilweise von der Perspektive des Betrachters abhängig ist und sich über die Zeit verändern kann. Dies lässt sich anhand des folgenden Beispiels erklären:

Johannes Keplers Hauptwerk *Astronomia nova* mag historisch einen wichtigen Beitrag zur Entwicklung der Astronomie geleistet haben. Durch seine Theorie zur Planetenbewegung wurde es erstmals möglich, die Bewegung von Himmelskörpern genauer zu bestimmen. Allerdings behandelt seine Arbeit auch Schlussfolgerungen, die unter heutiger Betrachtung als falsch gewertet würden. Für den Historiker hingegen stellt das Werk auch heute eine wichtige Quelle dar. Astronom und Historiker würden heute bei der Bewertung der Arbeit also wahrscheinlich zu zwei unterschiedlichen Urteilen gelangen.

Um allgemein die inhaltliche Qualität einer wissenschaftlichen Arbeit messen zu können, müssen zuvor Indikatoren ermittelt werden, die das Repräsentationstheorem erfüllen. Es sprengt den Rahmen dieser Arbeit, würde hier der aufwendige Nachweis geführt werden, an bestimmten Indikatoren dieses Theorem zu testen. Stattdessen soll auf allgemein anerkannte Indikatoren zurückgegriffen werden, wohl wissend, dass sie von Teilen der Fachwelt als problematisch eingestuft werden.

Eine Möglichkeit der Auswahl geeigneter Indikatoren besteht darin, sich die Anweisungen für Peers in einflussreichen Zeitschriften anzuschauen. Im Folgenden wurde dies für die fünf bestplatzierten englischsprachigen Zeitschriften bei *Google Scholar Metrics* getan. Dahinter steht die Überlegung, dass diese Zeitschriften u. a. deshalb so erfolgreich sind, weil die Peers qualitativ hochwertige Arbeit leisten, die wiederum auf den speziellen Empfehlungen der Redaktionen basiert. **Tabelle 3** gibt einen Überblick über die Empfehlungen.

Journal	Empfohlene Bewertungskriterien
Nature	<i>Provides strong evidence for its conclusions. Novel (we do not consider meeting report abstracts and preprints on community servers to compromise novelty). Of extreme importance to scientists in the specific field. Ideally, interesting to researchers in other related disciplines. In general, to be acceptable, a paper should represent an advance in understanding likely to influence thinking in the field. There should be a discernible reason why the work deserves the visibility of publication in a Nature journal rather than the best of the specialist journals.</i>
The New England Journal of Medicine	<i>Nicht öffentlich</i>
Science	<i>Most influential in their fields significantly advance scientific understanding novel data, syntheses or concepts broadly important data, syntheses or concepts should merit the recognition by the scientific community</i>
The Lancet	<i>studies that will appeal to a wide general readership. Also, the question being addressed and the planned design and analysis will need to be as original as possible, topical, and valid.</i>
Cell	<i>objective assessment of both the technical rigor and the novelty of the presented work</i>

Tabelle 3: Zusammengestellte Bewertungskriterien für Reviewers der fünf bestplatzierten Journals nach GSM

Aus den hier genannten spezifischen Kriterien, können allgemeine Kriterien (bzw. Kriterien-gruppen) hergeleitet werden

- Hoher Neuigkeitswert / Originalität
- Belastbare Beweise / Gültigkeit / Nachvollziehbarkeit
- Von Bedeutung für das entsprechende Fachgebiet
- Von Interesse für andere Fachgebiete
- Nachhaltig für den wissenschaftlichen Fortschritt

Die beiden Kriterien hinsichtlich der Bedeutung für das eigene Fachgebiet bzw. für andere Fachgebiete lassen sich als Effektivitätsproblem der jeweiligen Journals betrachten. Das Ziel, einen hohen Impact-Wert zu erhalten, lässt sich nur dann erreichen, wenn die veröffentlichten Artikel einen möglichst breiten Leserkreis finden. Dies gilt im eingeschränkten Maße auch für spezielle Fachzeitschriften. Das Problem stellt sich für Betreiber von OA-Repositorien nicht in dieser Form, weshalb im Rahmen dieser Arbeit diese beiden Kriterien nicht weiter verfolgt werden sollen.

Von den verbleibenden Kriterien ist keines unproblematisch.

Dies betrifft insbesondere das Kriterium der **Originalität**. Ein kürzlich veröffentlichter Beitrag kann heute durch die verwendete Methode oder aufgrund der enthaltenen Kernaussagen originell erscheinen; zukünftige Leser werden diesen Eindruck wahrscheinlich nicht mehr teilen können. Um als originell zu gelten, muss der Zeitpunkt der Bewertung vermerkt werden. Bei Zeitschriften ist dies durch das Erscheinungsdatum und den vorausgegangenen Peer-Review-Prozess gegeben. Letztendlich ist es schwierig, den Leser eindeutig zu instruieren, ob er bei der Bewertung von Originalität das Veröffentlichungsdatum oder das Datum seiner Bewertung verwenden soll. Sollte sich das Bewertungsdatum durchsetzen, ergibt sich jedoch daraus auch die Chance, die zeitliche Entwicklung der „Originalität“ einer Arbeit zu erfassen und auszuwerten. OA_SCORE wird dadurch um einen Aspekt erweitert, der über das einfache Messen der inhaltlichen Qualität hinausgeht.

Das Kriterium der Validität kann über OA_SCORE nicht gemessen werden. Auch dies hängt mit dem beabsichtigten Zeitpunkt der Bewertung zusammen. Das Instrument ist dafür vorgesehen, dass der Leser unmittelbar nach der Lektüre seine Eindrücke zusammenfasst. Unter dieser Voraussetzung sind in vielen Fachgebieten, wie z. B. die Hochenergiephysik, kaum fundierte Validitätsprüfungen zu erwarten. Allerdings bietet es sich an, nach der Plausibilität der Beiträge zu fragen.

Unter **Plausibilität** soll hier die Übertragung von Validität auf eine subjektive Ebene verstanden werden. Der Leser als Subjekt bewertet seinen persönlichen Eindruck, den er – befragt nach der Validität einer Arbeit – gewonnen hat. Plausibilität hat dann selbstverständlich nicht mehr die gleiche Aussagekraft wie Validität, da sie durch Faktoren wie z. B. die äußere Form der Arbeit auch suggeriert werden kann.

Ich setze trotzdem voraus, dass es einen Zusammenhang zwischen Plausibilität und Validität gibt, der umso deutlicher wird, je mehr Bewertungen abgegeben werden.

Auch Angaben zur Plausibilität sind abhängig vom Zeitpunkt der Bewertung. Neue Erkenntnisse können vorangegangene Arbeiten plausibler oder aber obsolet machen. Ein klassisches Beispiel für die erstgenannte Entwicklung wäre die Theorie Alfred Wegeners zur Kontinentaldrift. Sie wurde zunächst von führenden Geologen für nicht plausibel gehalten, bevor Jahre später wissenschaftlichen Beobachtungen die wichtigsten Aussagen dieser Theorie bestätigten (**Willmann 2012**).

Das von den Zeitschriftenverlagen erhobene Kriterium der wissenschaftlichen Nachhaltigkeit deutet auf die Messung des zu erwarteten *Impacts* hin. Es ist verständlich, wenn Zeitschriftenverlage von den Gutachtern verlangen, nur solche Beiträge positiv zu bewerten, die eine größtmögliche Anzahl an Zitationen versprechen. Die Expertise der Gutachter besteht darin, diesen *Impact* vorausszusehen. Übertragen auf OA_SCORE bedeutet dies, dass der Leser bewerten soll, welchen Einfluss der soeben gesichtete Beitrag auf seine eigene Arbeit hat. Diese relevante Qualität ließe sich auf der persönlichen Ebene als **Nützlichkeit** übersetzen. Es ist abzusehen, dass die unkommentierte Verwendung des Kriteriums Nützlichkeit in OA_SCORE zu mehrdeutigen Ergebnissen führen wird. Die Einschätzung von Nützlichkeit geschieht immer in einem bestimmten Zusammenhang. Im Fall von OA_SCORE sind zwei Beziehungen zu unterscheiden. Einerseits die Beziehung der relevanten Qualität des zu bewertenden Beitrags zum Informationsbedürfnis des Wertenden, andererseits die Beziehung der relevanten Qualität zu einem bestimmten Fachgebiet. Erstere ist subjektiv angelegt und kann deshalb für OA_SCORE nur benutzt werden, wenn zugleich zusätzliche Informationen über die Arbeiten des Bewertenden vorliegen. Die zweite Beziehung setzt eine allgemein gültige Struktur von Fachgebieten sowie deren korrekte Einordnung voraus. Alternativ ließe sich die Zugehörigkeit einer Arbeit zu einem Fachgebiet über die semantische Verknüpfung von Stichwörtern ermitteln.

Für die vorliegende Untersuchung soll dieses Verständnis von Nützlichkeit aus Gründen des einfacheren Gebrauchs trotzdem Verwendung finden. Um das zu Grunde gelegte Informationsbedürfnis zu standardisieren, wird im Rahmen der Befragung ein zusätzliches Kriterium als Vergleichsmaßstab eingefügt.

Zu den drei aufgestellten Kriterien Originalität, Plausibilität und Nützlichkeit müssen nunmehr noch folgende Fragen beantwortet werden:

- Gibt es einen prägnanten Begriff, womit sich das Konzept des Kriteriums (bzw. des Indikators) beschreiben lässt?
- Lässt sich die zuvor genannte Frage auch für die anderen weit verbreiteten Sprachen in OA-Repositorien lösen? Vierkant zufolge wären dies - nach Englisch - in absteigender Häufigkeit: Französisch, Spanisch, Deutsch, Japanisch, Polnisch, Chinesisch, Portugiesisch, Niederländisch und Italienisch (**Vierkant 2012, 60**). Im Rahmen dieser Arbeit sollen Englisch, Französisch, Spanisch und Deutsch untersucht werden, beim Englischen zudem die beiden Varianten *British English* und *American English*.
- Gibt es gegenseitige Abhängigkeiten zwischen den Kriterien/Indikatoren?

Zur Beantwortung dieser Frage sollen die Begriffe Originalität, Plausibilität und Nützlichkeit in Bezug auf die Häufigkeit ihres Auftretens, ihre Definitionen sowie möglicher Synonyme untersucht werden. Gleiches erfolgt mit den treffendsten Synonymen. Grundsätzlich sollen dabei Nomen bevorzugt werden. Um die Popularität der Begriffe zu ermitteln, wird der **Ngram Viewer** von Google verwendet und die einfache mittlere Häufigkeit des Auftretens des Begriffes im Korpus von Google Books in den Jahren 2000 - 2008 festgestellt. Dem liegt die Annahme zu Grunde, dass populäre Begriffe - sofern sie zur akzeptierten Schriftsprache gehören - häufiger veröffentlicht werden. Da der *Ngram Viewer* zwischen Groß- und Kleinschreibung unterscheidet, werden jeweils die Werte für die komplett klein geschriebenen Begriffe bzw. die Begriffe mit Großbuchstaben am Wortanfang übernommen und addiert. Die Häufigkeit von komplett in Versalien gefassten Begriffen wird ignoriert.

Eine detaillierte Übersicht der ermittelten Werte findet sich im **Anhang 10.1**.

Für die weitere Untersuchung werden jeweils die beiden Begriffe mit der größten Häufigkeit ausgewählt, es sei denn, dass deren Definition zu viele Mehrdeutigkeiten aufweisen oder nicht auf den Sachverhalt zutreffen (**Tabelle 4**).

EN	FR	ES	DE
Novelty Newness	Unicité Originalité	Originalidad Singularidad	Originalität Einfallsreichtum
Credibility Plausibility	Crédibilité Fiabilité	Credibilidad Plausibilidad	Glaubwürdigkeit Plausibilität
Relevance Usefulness	Utilité Valeur	Utilidad Relevancia	Relevanz Nützlichkeit

Tabelle 4: Ermittelte Begriffe und Synonyme für die zu bewertenden Kategorien in vier Sprachen

Wie zu sehen ist, verfügen im Deutschen Glaubwürdigkeit und Relevanz über eine höhere Popularität als Plausibilität und Nützlichkeit. Sie sollen deshalb im weiteren Verlauf der Arbeit als Beschreibung der Kriterien dienen.

Die Alternative *valeur* im Französischen steht eigentlich wegen der Vieldeutigkeit nicht zur Diskussion, wird jedoch mangels weiterer Alternativen hier mit aufgenommen.

Bei der zu erstellenden Umfrage wird es nach dem bisher Gesagten u. a. darauf ankommen, signifikante Präferenzen der Befragten für eine der beiden Alternativen je Kriterium zu ermitteln.

Es ist vorgesehen, die Präferenz zu Beginn des Interviews abzufragen. Während des weiteren Interviews darf sie einmal verändert werden. Am Schluss des Interviews wird die Präferenz nochmals abgefragt, um deren Stabilität zu überprüfen. Da die Befragten sehr wahrscheinlich dazu tendieren werden, ihre zuvor gefasste Meinung beizubehalten, käme einem eventuellen Meinungswechsel eine besondere Aussagekraft zu.

Neben der Präferenz für einen Begriff, sollen auch jene zwei Kategorien ausgewählt werden, die für eine zweidimensionale Skala in Betracht kommen. Auch hier erhalten die Befragten die Möglichkeit, ihre Wahl im Verlauf des Gesprächs zu ändern und am Ende zu bekräftigen bzw. abzuschwächen.

In der zweiten Phase des Interviews sollen die Befragten anhand der von ihnen gewählten Kriterien und Begriffe verschiedene Texte bewerten. Dazu werden ihnen die passenden Skalen und dazugehörigen Erläuterungen vorgelegt. Da die Interviewzeit knapp bemessen ist, handelt es sich bei den Texten um keine vollständigen wissenschaftlichen Veröffentlichungen, sondern um Kurztex te mit maximal zwei Absätzen. Es wird davon ausgegangen, dass die Texte trotz deren Kürze eine Bewertung zulassen.

Der Inhalt der Texte soll so justiert sein, dass dadurch eindeutige Bewertungen ermöglicht werden. Um einen Maßstab für Relevanz zu geben, wird zuvor ein Thema vorgegeben, zu welchem relevante Literatur herausgesucht werden soll. Für die Befragung wurde hier das Thema „Kann Afrika als Wiege der Mathematik gelten?“ ausgewählt. **Tabelle 5** zeigt eine Übersicht der ausgewählten Texte für den deutschen Fragebogen.

Während des Gesprächs soll vermerkt werden, wie die Befragten die schriftlichen Hinweise zu den Skalen nutzen.

Bei der Bewertung des zweiten Textes soll die postulierte Abhängigkeit der Kriterien Glaubwürdigkeit und Relevanz näher untersucht werden. Eine Veröffentlichung, die keine Glaubwürdigkeit besitzt, sollte üblicherweise auch nicht relevant sein. Das bedeutet, dass bei Verwendung einer 1-von-8-Skala in der Reihenfolge Originalität, Glaubwürdigkeit und Relevanz die Aussagen 1-0-1 und 0-0-1 eigentlich nicht benötigt werden. Analog dazu sollte bei einer

zweidimensionalen Skala unter Verwendung dieser beiden Kriterien nur der untere linke Quadrant ausgewählt werden.

Die Umfrage soll, um den Rahmen dieser Arbeit nicht zu sprengen, nur mit Französisch- und Deutschsprechern durchgeführt werden. Die meisten Umfragen sollen in Kamerun durchgeführt werden. Hierbei kommt erschwerend hinzu, dass viele kamerunische Teilnehmer wahrscheinlich doppelt halbsprachig sein werden. Für diese Personen gibt es keine exklusive Muttersprache, die sie perfekt beherrschen. Folglich können Schwierigkeiten beim Verständnis der Aufgabenstellung erwartet werden.

Die gewählten Kriterien sind dann für die Nutzung in OA_SCORE geeignet, wenn folgende signifikanten Beziehungen deutlich werden:

- Die Gesamtheit der pro Text abgegebenen Bewertungen entspricht den zu erwartenden Ergebnissen, wobei bei den Texten 1-3 lediglich die Abwesenheit von jeweils Originalität, Glaubwürdigkeit bzw. Relevanz untersucht werden soll.
- Die Mehrzahl der Befragten entscheidet sich für ein bestimmtes Paar von Kriterien für die zweidimensionale Skala.
- Die Mehrzahl der Befragten entscheidet sich für eine bestimmte Kombination von Begriffen in der jeweiligen Sprache

Zudem kann eine Skala als besonders geeignet angesehen werden, wenn ein ausreichender Anteil der Befragten diese für die geeignetste Skala hält und die genannte Skala zugleich plausible Ergebnisse liefert.

Die beiden Fragebögen sind im **Anhang 10.5** wiedergegeben.

5.3.2 Pretest

Ein Pretest mit deutschen Muttersprachlern und n=5 führte zu folgenden Beobachtungen:

- Eine Kurzanleitung zur Benutzung der Skalen muss unmittelbar neben der jeweiligen Skala angegeben sein. Der Wegweiser zu einer Hilfeseite wurde in 4 von 5 Fällen nicht wahrgenommen.
- Die Abhängigkeit der Relevanz von der Plausibilität und umgekehrt wird anders wahrgenommen als vermutet. So lautete eine sinngemäße Aussage zu Text 3: „Wenn es für mich nicht relevant ist, brauche ich auch nicht nach der Plausibilität fragen.“
- Die Wahl der verwendeten Begriffe je Kriterium spielt nur eine untergeordnete Rolle.
- Alle Befragten kamen im Verlaufe der Zeit immer besser mit der Aufgabe klar. Es scheint, dass sich die Zeitspanne zur Beantwortung einer Frage immer mehr verkürzt hat.
- Alle Befragten gaben einer der beiden dreidimensionalen Skalen den Vorzug, obgleich sie gleichzeitig zugaben mit diesen Skalen Schwierigkeiten gehabt zu haben.
- Text 4 wird von Fachfremden – entgegen der Vermutung – nicht immer als Nonsens aufgefasst.
- Der Hinweis auf die Wikipedia als Quelle für Text 3 scheint zu bewirken, dass der Text als weniger plausibel betrachtet wird.

Aus diesen Beobachtungen ergeben sich folgende Anpassungen:

- Der Hinweis auf die Hilfeseite muss grundsätzlich erfolgen, bevor der erste Text bewertet wird, es sei denn der Befragte findet diesen Hinweis selbst

- Es sollte zusätzlich erfasst werden, wie viel Zeit für eine Bewertung notwendig ist. Hieraus können sich Hinweise für einen Lerneffekt ergeben, vorausgesetzt die Güte der Antwort nimmt im Verlauf der Befragung nicht ab.
- Der Hinweis auf die Wikipedia in Text 3 wird durch eine erfundene Quellenangabe ersetzt.
- Es muss bei der Besprechung der Ergebnisse berücksichtigt werden, dass Plausibilität bei nicht vorhandener Relevanz tendenziell schlechter bewertet wird, sowie dass der Text 4 nicht immer als Nonsens aufgefasst wird.

Text 1	<p>Enthält keine Originalität.</p> <p>„... Im alten Ägypten wurden sehr große Pyramiden gebaut. Pyramiden sind geometrische Körper mit einem Vieleck als Grundfläche. Das unterscheidet sie vom Kegel, der eine runde Grundfläche hat. Wir wissen leider nicht ganz genau, warum sich die alten Ägypter anstatt für die Kegelform für eine Pyramidenform entschieden haben. In jeden Fall war der Bau der Pyramiden eine beachtliche Leistung, die ohne grundlegende Kenntnisse in der Mathematik nicht zu bewerkstelligen ist. ...“ (http://pyramiden-mathematik.edu) {Quelle erfunden}</p>
Text 2	<p>Enthält keine plausible Aussage, wäre jedoch für das Thema relevant</p> <p>„...Der Neger stellt, wie schon gesagt worden ist, den natürlichen Menschen in seiner ganzen Wildheit und Unbändigkeit dar: von aller Ehrfurcht und Sittlichkeit, von dem, was Gefühl heißt, muß man abstrahieren, wenn man ihn richtig auffassen will; es ist nichts an das Menschliche Anklingende in diesem Charakter zu finden. Die weitläufigen Berichte der Missionare bestätigen dieses vollkommen, und nur der Mohammedanismus scheint das einzige zu sein, was die Neger noch einigermaßen der Bildung annähert...“ (Hegel: Vorlesungen über die Philosophie der Geschichte)</p>
Text 3	<p>Enthält keine relevante Aussage</p> <p>„Die Faultiere (Folivora) bilden eine urtümliche Unterordnung der zahnarmen Säugetiere (Pilosa) und sind mit Ameisenbären und Gürteltieren verwandt. Es sind sechs rezente Arten bekannt, die sich in die beiden Familien der Zweifinger-Faultiere (Megalonychidae) und der Dreifinger-Faultiere (Bradypodidae) aufteilen. Daneben gab es noch eine Reihe heute ausgestorbener Riesenfaultiere. Faultiere bewohnen hauptsächlich die Baumkronen der tropischen Regenwälder von Mittelamerika und dem Amazonasbecken bis zum südlichen Brasilien. ...“ (www.wikipedia.org)</p>
Text 4	<p>Ist ein Nonsens-Text, enthält also weder Originalität, noch Glaubwürdigkeit noch Relevanz</p> <p>"... Wir haben in diesem Papier gezeigt, dass Moores Gesetz autonom, elektronisch und verlustfrei gemacht ist und ROE keine Ausnahme von dieser Regel darstellt. Wir verwendeten dezentralen Archetypen um zu beweisen, dass Model-Checking und Suffix-Bäume kontinuierlich unvereinbar sind. Wir verwendeten modularen Methoden um zu argumentieren, dass lokale Netzwerke und DHCP keine Verbindung zu diesem Prinzip eingehen. Schließlich haben wir einen Rahmen für Moores Law (ROE) codiert, bestreitend, dass Architektur und SMP in der Regel nicht kompatibel sind.“ (Miller J, DuBois M: Studying SCSI Disks and Superblocks . Journal for Scientific Generators, Vol 2 (2005), p. 23-27) {Quelle erfunden}</p>
Text 5	<p>Enthält ein hohes Maß an Originalität und Glaubwürdigkeit und ist für das vorgegebene Thema relevant</p> <p>„Africa's Homo Sapiens Were the First Techies Dec. 5, 2012 — The search for the origin of modern human behaviour and technological advancement among our ancestors in southern Africa some 70,000 years ago has taken a step closer to firmly establishing Africa, and especially South Africa, as the primary centre for the early development of human behaviour. A new research paper by Wits University archaeologist, Prof. Christopher Henshilwood, is the first detailed summary of the time periods he and a group of international researchers have been studying in South Africa: namely the Still Bay techno-traditions (c. 75,000 -- 70,000 years) and the Howiesons Poort techno-tradition (c. 65,000 -- 60,000 years). The paper, entitled 'Late Pleistocene Techno-traditions in Southern Africa: A Review of the Still Bay and Howiesons Poort, c. 75 ka,' has been published online in the Journal of World Prehistory on Nov. 6, 2012. ...“ (http://www.sciencedaily.com/releases/2012/12/121205103005.htm. Abgerufen am 18.2.2013)</p>

Tabelle 5: Charakteristik der zur Befragung verwendeten Texte

6 Untersuchungen

6.1 Metadaten-Abfrage von OA-Repositoryn

6.1.1 Datenkorpus

Für die Abfrage der OA-Repositoryn kam das bereits erwähnte Skript zur Anwendung. Die Abfrage erfolgte mit Unterbrechungen in der Zeit vom 3.2.2013 bis 18.4.2013 mit einer Hauptphase vom 3.2.2013 bis 7.4.2013

Es wurden die verfügbaren Informationen zu OAI-PMH-Schnittstellen aus den beiden Meta-Repositoryn *ROAR* und *OpenDOAR* ausgewertet. Es stellte sich schnell heraus, dass nur ein Teil der darin aufgelisteten Repositoryn über die notwendige Angabe zur OAI-PMH-Schnittstelle verfügte. Zudem listet das *ROAR* auch Repositoryn auf, bei denen das Element `<open_access>` den Wert *FALSE* hat. **Tabelle 6** gibt einen Überblick über die festgestellten Mengen an Repositoryn. Die Werte wurden teilweise über BASH-Kommandos ermittelt. (**Anhang 10.3**)

	ROAR	OpenDOAR	Bezeichnung der Menge
<i>Datum der Abfrage</i>	22.01.2013	26.01.2013	
<i>Enthaltene Datensätze</i>	3.343	2.254	
<i>Davon als Open Access deklariert</i>	1.397	2.254	
<i>Davon mit Angabe der OAI-PMH-Schnittstellenadresse</i>	1.230	1.479	
<i>Vereinigungsmenge der Schnittstellenadressen ROAR \cup OpenDOAR</i>	2.230		M_v
<i>Davon waren per HTTP-Request erreichbar</i>	2.077		M_E
<i>Davon lieferten eine Aussage über vorhandenen Metadatenformate</i>	1.537		M_M
<i>Davon verfügten über mindestens einen Record bzw. lieferten einen validen XML-Stream bei der Datensatzabfrage</i>	1.453		M_R
<i>Davon konnten mittels METS abgefragt werden</i>	703 (732 von M_M)		M_{METS}
<i>Verbleibende Repositoryn, die mittels OAI_DC abgefragt wurden</i>	750 (805 von M_M)		M_{OAI_DC}
<i>Teilmenge von M_R bei der mindestens bei einem Record eine Formatangabe ermittelt werden konnte</i>	1.036		M_F

Tabelle 6: Übersicht über die Menge der untersuchten Repositoryn

An den Rohdaten mussten aus verschiedenen Gründen die folgenden Korrekturen angewandt werden:

- Der Abfrageprozess erfolgte in mehreren Phasen. Im Ergebnis entstanden mehrere Dateien mit Rohdaten, die zunächst zusammengeführt werden mussten.
- Durch einen Fehler im Skript wurde unter bestimmten Bedingungen beim Wechsel zum nächsten Repositorym der Zähler nicht zurückgesetzt. Der Fehler betraf aufeinander folgende Repositoryn mit nicht-validem XML bei der Datensatzabfrage. Entsprechend übernahmen die jeweiligen Repositoryn die Zahlen zu den Formatangaben vom Vorgänger. Die Zählvariablen in allen betroffenen Repositoryn wurden manuell auf Null gesetzt.
- Das Skript setzt den Zähler für die Formatangaben zurück, in dem es alle Zahlenwerte auf 0 setzt. Davon sind auch Zahlenwerte innerhalb von Format-Bezeichnungen

betroffen (z. B. *image/jp0* statt *image/jp2*). Dadurch wurden im Verlaufe der Abfrage einige wiederkehrende Formatangaben nicht korrekt zusammenaddiert. Die Korrektur der Bezeichnungen ist nachträglich erfolgt. Zur Ermittlung der Gesamtzahl des Vorkommens einer relevanten Formatangabe wurden die Zwischensummen addiert.

Die korrigierten Rohdaten sind über den **Anhang 10.6** verfügbar.

In einem weiteren Schritt wurde untersucht, inwieweit die gesammelten Daten verlässlich sind. Dabei wurden folgende mögliche Fehlerquellen berücksichtigt.

Unkontrollierter Abbruch der Abfrage. Es wird vermutet, dass in mehreren Fällen nicht das gesamte Repositorium abgefragt wurde. Dies kann durch einen Fehler in der Repositoriums-Software begründet sein, z. B. wenn vorzeitig ein leeres *ResumptionToken* - als Ende-Zeichen - ausgeliefert wurde. Ebenso ist denkbar, dass das zuvor ausgelieferte *ResumptionsToken* verfallen ist, während die zuvor abgerufenen Daten analysiert wurden. Es lässt sich an dieser Stelle nur vermuten, welcher Anteil der abgefragten Repositorien von diesem Abbruchtyp betroffen ist. Einen Hinweis gibt die festgestellte Anzahl der abgefragten Records. Nach Einsicht der umfangreichen Log-Dateien wurde abgeschätzt, dass in den meisten Fällen pro Abfrage 30, 50, 100, 250 oder 500 Records geliefert wurden. Das Häufigkeitsmaximum scheint bei 100 Records pro Abfrage zu liegen.

Durch die Ermittlung des jeweiligen Divisionsrests kann die Wahrscheinlichkeit für einen erfolgten Abbruch geschätzt werden. Dies ist z. B. denkbar wenn ein Repositorium genau 3.600 Records in Paketen zu jeweils 100 Records lieferte. Die Wahrscheinlichkeit, dass das Repositorium eine durch 100 teilbare Zahl von Records aufweist, beträgt bei zufälliger Verteilung $P=0,01$.

Die folgende Abbildung stellt dar, wie häufig jeweils ein Divisionsrest 0 für die fünf genannten Divisoren vorkommt (**Anhang 10.8**).

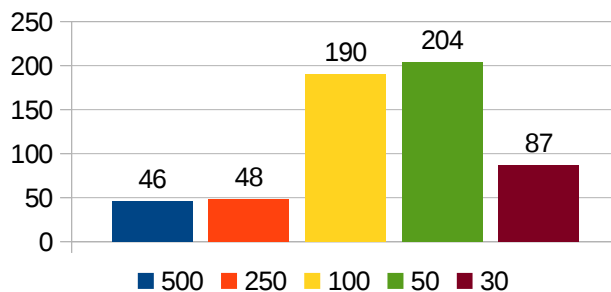


Abbildung 6: Anteil der Repositorien ($n=M_F=1036$), bei der die Gesamtzahl der abgefragten Records ein Vielfaches von bestimmten Werten ist.

Da die genaue Verteilung der Zahl der Records je Abfrage nicht bekannt ist, können hier nur die folgenden, sehr groben Aussagen gemacht werden.

Es gibt in der Tat Repositorien, die nicht vollständig abgefragt wurden. Deren Anzahl beträgt im ungünstigsten Fall etwa 26 %. Dem liegt die Überlegung zu Grunde, dass 500, 250 und 100 jeweils Vielfache von 50 sind und ein Fünftel der Vielfachen von 30 zugleich auch Vielfache von 50 sind. Es wäre nun denkbar, dass genau alle Repositorien mit einer Anzahl an Records, welche einem Vielfachen von 50 entspricht, zufällig auch in 50er Schritten abgefragt wurden. Gleiches gilt für Repositorien, deren Zahl an Records ein Vielfaches von 30 entspricht und die zufälligerweise alle in 30er Schritten abgefragt wurden.

$$\frac{(204 + \frac{4}{5} \cdot 87)}{1036} \approx 0,26 \quad (1)$$

Umgekehrt gibt es im günstigsten Fall etwa 8 % an nicht vollständig abgefragten Repositorien, wenn man von der Möglichkeit absieht, dass die Anzahl der Datensätze eines Repositoriums bewusst oder zufällig genau ein Vielfaches der genannten Werte ist.

$$\frac{\left(\frac{1}{5} \cdot 87 + \frac{1}{3} \cdot 204\right)}{1036} \approx 0,08 \quad (2)$$

Aus diesen beiden Extremen ergibt sich ein arithmetisches Mittel von etwa 16 % bzw. 166 unvollständig ausgewerteten Repositorien. Der Wert ist jedoch angesichts der fehlenden Informationen kaum belastbar.

Es wird im weiteren vereinfachend angenommen, dass die Formatangaben in den betroffenen Repositorien weitgehend zufällig verteilt sind, so dass im Bezug auf das Ziel der Untersuchung keine wesentliche Einschränkung besteht. In Wirklichkeit dürfte die Verteilung der Formatangaben innerhalb eines Repositoriums jedoch stark von der historischen Entwicklung des Repositoriums abhängen.

Eine weitere mögliche Fehlerquelle sind **Ausreißer**. Darunter sollen Repositorien verstanden werden, bei denen ein ungewöhnliches Verhältnis zwischen der Anzahl der Records mit Formatangaben und der Anzahl der Formatangaben selbst festgestellt wurde. So konnten 35 Repositorien festgestellt werden, bei denen dieses Verhältnis 11 oder mehr beträgt. Das heißt jeder Dokumenten-Record enthält durchschnittlich mehr als 11 Formatangaben. Dies kann unter bestimmten Bedingungen durchaus noch im Rahmen des Normalen liegen, z. B. wenn ein Record aus mehreren Seiten eines eingescannten Druckwerks besteht, bei dem jede Seite als einzelne Datei abgelegt ist. Ein wirklicher Ausreißer in diesem Bereich scheint nur ein Repository der TU-Darmstadt (tukart.ulb.tu-darmstadt.de) zu sein, bei dem ein Durchschnitt von 1.260 Formatangaben pro Record festgestellt wurde! Bei einer manuellen Überprüfung konnten die Records zwar über *OAI_DC* (ohne Formatangaben), aber nicht mehr mittels *METS* abgerufen werden. Genauer gesagt, es erfolgte keine Fehlermeldung und der Versuch wurde nach etwa 20 Minuten abgebrochen. Bei der Abfrage mittels BASH-Skript wurden 489 Dokumenten-Records mit 616.260 Formatangaben ermittelt, darunter keine Angabe mit MIME-Typ *application/pdf*, jedoch 613.205 Angaben mit *image/jpeg*. Die Zahl von 500 abgerufenen Records lässt zudem einen Abbruch vermuten.

Da es sich aber laut Betreiber um ein Repository mit alten eingescannten Drucken handelt, wurde das Ergebnis für plausibel genug befunden, um in der Datenmenge zu verbleiben. Zudem spricht es für die Nullhypothese.

Auf der anderen Seite gibt es ein Repository, welches scheinbar mehr Records mit Formatangaben führt, als Formatangaben selbst ermittelt werden konnten. (<http://lib-repos.fun.ac.jp/dspace-oai/request?verb=Identify>). Warum dies so ist, kann nicht abschließend geklärt werden. Es wird deshalb davon ausgegangen, dass beim Zählen der Dokumenten-Records ein Fehler aufgetreten ist. Die Angaben zur Verteilung der einzelnen Formate wären dem entsprechend korrekt, weshalb auch dieses Repository im Datenbestand verbleiben soll. Gezählt wurden hier 396 Angaben *application/pdf* sowie 447 Angaben *text/plain*.

Insgesamt wurden 201 verschiedene Metadatenformate ermittelt. Davon werden 41 in jeweils 10 oder mehr Repositorien verwendet. **Abbildung 7** zeigt die Häufigkeitsverteilung der Metadatenformate mit $n \geq 10$

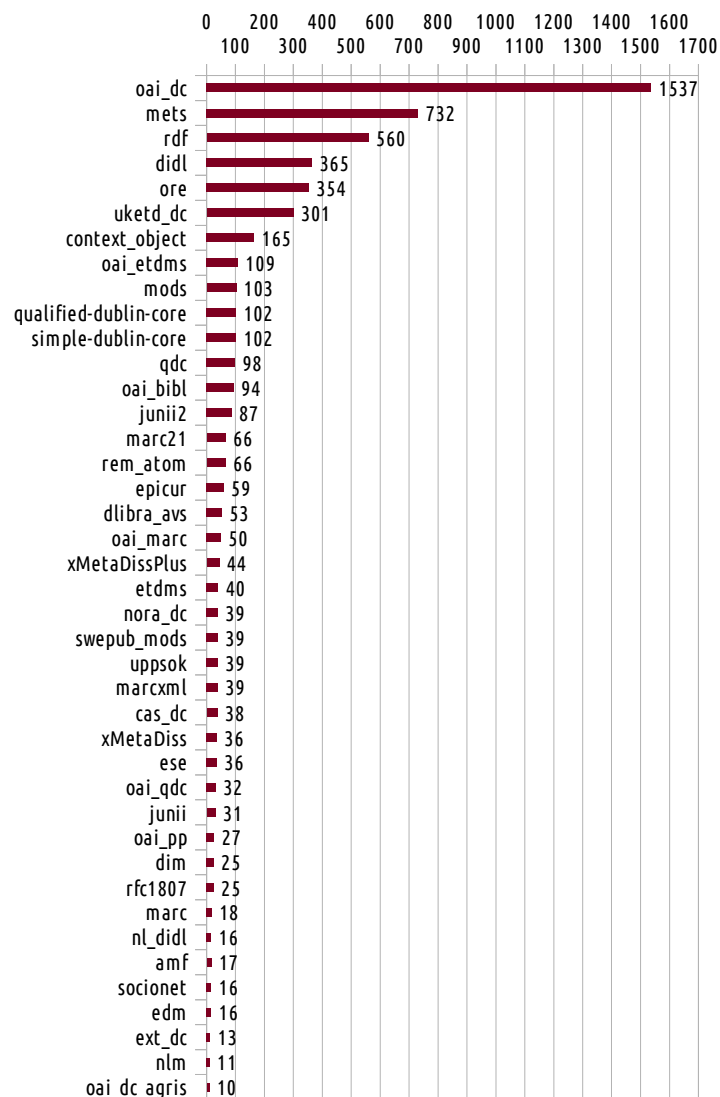


Abbildung 7: Häufigkeitsverteilung von Metadatenformaten in den untersuchten Repositorien (abgebildet sind nur die am häufigsten vorkommenden Formate)

Erwartungsgemäß sind die beiden für die Untersuchung herangezogenen Metadatenformate *OAI_DC* bzw. *METS* am häufigsten vertreten. Es kann nun sein, dass die Analyse mittels eines der beiden nachfolgend platzierten Metadatenformate (*RDF* bzw. *DIDL*) entscheidende Vorteile im Sinne der Aufgabenstellung gebracht hätte. Dazu wurde untersucht, wie viele Repositorien zwar über eine dieser beiden Metadatenformate, nicht jedoch über *METS* verfügen. Die so identifizierten Repositorien wurden weiterhin darauf getestet, ob bereits über die Abfrage mit *OAI_DC* Formatangaben gewonnen wurden bzw. ob bei den dann übrig blei-

benden Repositorien überhaupt Records ermittelt wurden. **Tabelle 7** gibt einen Überblick über die gefundenen Ergebnisse, wobei als Ausgangsmenge $M_M=1.536$ dient.

Wie zu erkennen ist, gibt es insgesamt 28 Repositorien, die eventuell Formatangaben liefern können, wenn die Abfrage über *DIDL* bzw. *RDF* erfolgt wäre. Dies stellt bezogen auf die Menge $M_R=1.453$ einen Fehler von etwa 1,8 % dar. Es ist nicht klar, wie viele dieser 28 Repositorien durch Verwendung der beiden alternativen Metadatenformate überhaupt Formatangaben liefern würden. Der Aufwand, dies zu ermitteln, scheint jedoch angesichts der Aufgabenstellung zu hoch.

		<i>RDF</i>	<i>DIDL</i>
1	Anzahl N	560	365
2	Vereinigungsmenge $N_{RDF} \cup N_{DIDL}$	711	
3	Davon zugleich ohne <i>METS</i>	54	
4	Davon wurden bereits Formatangaben durch <i>OAI_DC</i> ermittelt	24	
5	Differenz (Zeile 3 - Zeile 4)	30	
6	Davon enthalten Datensätze	28	

Tabelle 7: Repositorien, deren Formatangaben möglicherweise mittels *RDF* bzw. *DIDL* identifiziert werden könnten

Von den über *OAI-PMH* abfragbaren Repositorien lieferten 1.036 Aussagen über Formate. (Menge: M_F). Darunter waren 672 Repositorien, die mittels *METS* abgefragt wurden (Menge: M_{F_METS}) sowie 364 Repositorien, die mittels *OAI_DC* abgefragt wurden (Menge: $M_{F_OAI_DC}$). **Tabelle 8** zeigt eine Übersicht zur Anzahl der ermittelten Formatangaben.

	<i>METS</i>	<i>OAI_DC</i>	<i>SUMME</i>
Anzahl Repositorien	672	364	1.036
Datensätze insgesamt	4.624.713	4.276.827	8.901.540
Records mit Formatangaben	3.175.866	1.935.536	5.111.402
Zahl der ermittelten Formatangaben	14.914.560	2.968.195	17.882.755

Tabelle 8: Übersicht über die Verteilung von Formatangaben

Erwartungsgemäß konnten die meisten Informationen zu Formaten durch Abfragen im Format *METS* gewonnen werden.

In einem nächsten Schritt soll die durchschnittliche Zahl der Formatangaben je Record mit nachgewiesener Formatangabe ermittelt werden. Der Einfachheit halber soll dieses Verhältnis im weiteren Verlauf der Arbeit als **Format-Record-Verhältnis** bezeichnet werden. Ausgangsmenge ist hierbei $M_F=1.036$. In **Abbildung 8** sind die Verhältnisse auf einer logarithmischen geordneten y-Achse dargestellt.

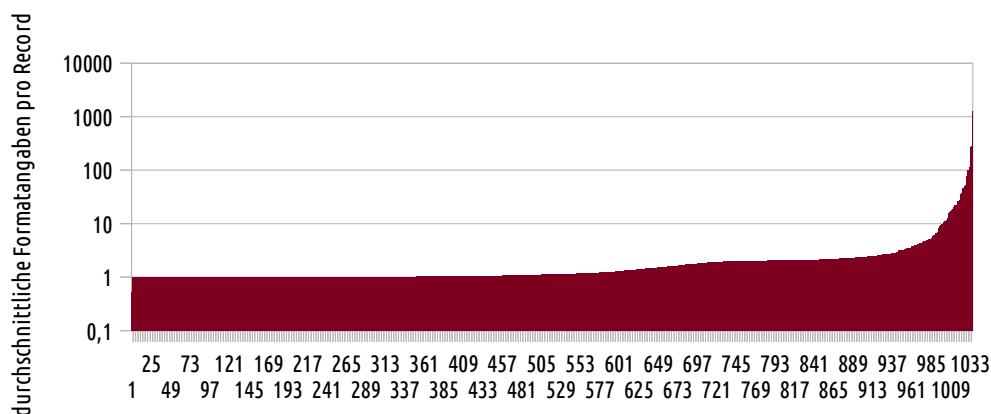


Abbildung 8: Verteilung des Verhältnisses der Formatangaben je Record (Format-Record-Verhältnis)

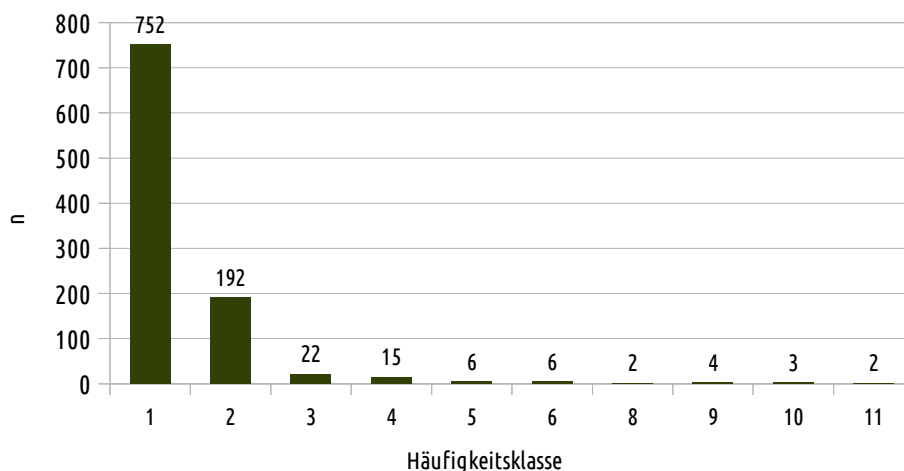


Abbildung 9: Verteilung des Verhältnisses der Anzahl von Records mit Formatangaben und der Anzahl der Formatangaben selbst, dargestellt in den 10 wichtigsten Häufigkeitsklassen

Zusätzlich wurden die ermittelten Verhältnisse Häufigkeitsklassen zugeordnet, wobei die Nummer der Klasse mit dem ganzzahligen Anteil des Format-Record-Verhältnisses übereinstimmt. Die zehn am häufigsten belegten Klassen sind in **Abbildung 9** dargestellt. Repositorien mit einem hohen Format-Record-Verhältnis lassen vermuten, dass dort vor allem historische Dokumente verwaltet werden, bei denen für jede Seite eine Datei angelegt wurde.

6.1.2 Ergebnisse

Es konnten in den untersuchten Datensätzen insgesamt 413 verschiedene Formatangaben ermittelt werden. Davon entsprechen 67 Angaben den offiziell von IANA deklarierten MIME-Typen. Die übrigen Formatangaben lassen sich den folgenden Gruppen zuordnen:

Experimentelle und nicht-standardisierte MIME-Typen. Diese sind entsprechend RFC 2045 mit einem „x-“ im Subtype gekennzeichnet. Entsprechend RFC 6648 wird diese Praxis heute nicht mehr empfohlen (**Borenstein et al. 1996, Nottingham et al. 2012**). Interessanterweise ist mit *image/x.djvu* ein solcher MIME-Typ mit Abstand am häufigsten gezählt worden. Der Grund für die Nichteinhaltung des Standards dürften die Anweisungen der Entwicklerfirma hinsichtlich der MIME-Typen-Deklaration des Formats gegenüber verschiedenen HTTP-Servern sein. (www.djvuzone.org, abgerufen am 19.4.13)

MIME-Typen-Angaben mit Tippfehlern. Offenbar verlangen einige Softwarelösungen die manuelle Eingabe des MIME-Typs bei der Aufnahme eines Items. Daraus ergeben sich zufällige Fehler. Ein Beispiel ist das Repositorium der *Universidad Autónoma de Madrid* (<http://digitool-uam.greendata.es/OAI-PUB>). Hier wurden insgesamt 14 verschiedene Schreibweisen des MIME-Typs *application/pdf* ermittelt. Bei 13 davon handelt es sich mit hoher Wahrscheinlichkeit um Tippfehler.

Pseudo-MIME-Typen. Diese Angaben ahmen standardisierte MIME-Typen nach, ohne die entsprechenden Standards zu berücksichtigen. Dabei ist nicht klar, ob es sich um zufällige Fehler bei der Eingabe von Items oder um systematische Fehler innerhalb der Softwarekonfiguration handelt. Ein Beispiel hierfür ist die Angabe *image/jpg*, welche sogar 10.002-mal ermittelt wurde. Die korrekte Angabe für JPEG-Dokumente lautet jedoch *image/jpeg* !

Formatangaben, die sich auf das repräsentierte Objekt beziehen. Hierzu gehören typische Angaben zur Beschreibung von physischen Medien, wie z. B. Seitenzahlen, Angaben zum Material (Papier, Pergament, usw.), Größenangaben, usw. Wie bereits erwähnt, erlaubt *Dublin Core* medienbezogene Angaben unter *<dc:format>*. Allerdings wurden derartige An-

gaben vereinzelt auch bei Abfragen mittels *METS* festgestellt. So findet sich zum Beispiel die Angabe *ocult* im Repository der *Universitat Pompeu Fabra*, Barcelona (<http://oai-repositori.upf.edu/request>). Eindeutig systematische Fehler stellen auch die knapp 1.200 Formatangaben *video* im Repository der *Universidade Federal de Paraná* dar (<http://dspace.c3sl.ufpr.br/dspace-oai/request>). Das Repository läuft unter DSpace und wurde mittels *METS* abgefragt.

Eine Übersicht zur Verteilung der häufigsten Formatangaben findet sich in **Abbildung 10**. Berücksichtigt sind hier nur Formatangaben, die mindestens 1.000-mal gezählt wurden. Auffallend ist die starke Schiefe der Verteilung. Z. B. wurden lediglich neun Formatangaben jeweils mehr als 100.000-mal gezählt.

Vier der in der Übersicht prominent vertretenden MIME-Typen müssen näher erläutert werden:

Image/x.djvu: Wie bereits erwähnt, handelt es sich um den häufigsten MIME-Typ in den untersuchten Repositorien. Der Anteil beträgt 44,4 % aller gezählten MIME-Typen! Es gibt einen starken Zusammenhang zwischen der Verwendung dieses MIME-Typs und der verwendeten Software. Alle 44 Repositorien, in denen die MIME-Type ermittelt wurde, laufen unter der Software *dLibra*. Innerhalb dieser Repositorien beträgt der Anteil des MIME-Typs gar 79,9 %. *dLibra* ist ein Produkt des *PSNC Digital Libraries Team* mit Sitz in Poznan. Die geographische Herkunft des Unternehmens dürfte auch der Grund dafür sein, dass ausnahmslos alle Repositorien mit dem MIME-Typ *image/x.djvu* die *Top Level Domain pl* führen. Die Häufigkeit des Auftretens dieses MIME-Typs erklärt sich aus dem überdurchschnittlich hohen Format-Record-Verhältnis, das für die genannten Repositorien im Durchschnitt etwa 35 beträgt!

Application/octet-stream: Es wird vermutet, dass in bis zu 77 Repositorien die Dokumente fälschlicherweise mit dem MIME-Typ *application/octet-stream* deklariert wurden. Laut **RFC 2046** wird dieser MIME-Typ jedoch nur für nicht weiter bestimmbar Byte-Streams empfohlen.

In 73 dieser 77 Repositorien wird ausschließlich der MIME-Type *application/octet-stream* angegeben. Insgesamt betrifft dies 392.997 Formatangaben. Stichproben in 10 der Repositorien ergaben jedoch, dass dort durchaus auch Dokumente eingestellt sind, die mit einem spezifischen MIME-Typ gekennzeichnet werden können. Die Stichprobe wurde dazu zufällig gezogen. Ein Repository war nicht mehr erreichbar. Alle anderen arbeiten auf der Basis der Software *EPrints*.

Im Rahmen dieser stichprobenartigen Untersuchung wurde zudem festgestellt, dass die Abfragen in allen 9 herangezogenen Repositorien unterschiedliche MIME-Typen bei unterschiedlichen Metadatenformat liefern. Konkret wurde festgestellt, dass bei einer Abfrage mittels *OAI DC* sehr wohl konkrete MIME-Typen (z. B. *application/pdf*) geliefert werden. Erfolgt die Abfrage jedoch über *METS*, liefern die gleichen Datensätze anstatt dessen *application/octet-stream* als MIME-Typ.

Im Rahmen dieser Arbeit können die Auswirkungen dieses systematischen Fehlers nicht ausreichend untersucht werden. Allerdings wird einmal mehr die Fehleranfälligkeit des Systems der maschinellen Metadaten-Abfragen mittels *OAI-PMH* in der Version 2.0 deutlich.

Obgleich der Fehler im Zusammenhang mit der verwendeten Software zu stehen scheint, sei erwähnt, dass es auch Repositorien auf der Basis von *EPrints* gibt, die bei beiden Metadatenformaten auch den gleichen MIME-Typ liefern.

Text/Plain: Einige Softwareanwendungen liefern zusätzlich zum Originaldokument noch eine einfache Textversion des Dokuments aus. Dies scheint zum Beispiel bei PDF-Dokumenten unter *DSpace* weitgehend der Fall zu sein. Eine Stichprobe mit $n=10$ unter *DSpace*-Repositorien bestätigt diese Annahme.

In der Tat ist der Anteil dieser MIME-Type in *DSpace*-Repositorien überdurchschnittlich hoch (Tabelle 9). Deshalb wird vermutet, dass ein hoher Anteil der Nennungen *text/plain* auf diese – weitgehend automatisch erzeugten – Hilfsdateien zurückgeht und es jeweils noch ein Dokument mit ursprünglichen Format im Record gibt.

	Total	DSpace	Anteil in %
Anzahl <i>text/plain</i>	1.421.193	531.146	37,4 %
Anzahl Repositorien (Menge M_F)	1.036	131	12,6 %

Tabelle 9: Anteil der MIME-Typen „text/plain“ in *DSpace*-Repositorien im Vergleich zur Gesamtzahl dieses MIME-Typ (Hinweis: Es wurden nur solche Repositorien gezählt, die das Wort „dspace“ in der URL führen.)

Text/html: Eine ähnliche Aussage lässt sich auch für den MIME-Typ *text/html* treffen. Es scheint in einigen Repositorien üblich, neben dem Original-Dokument noch eine HTML-Version auszuliefern. Dies trifft zum Beispiel auch für den EDOC-Server der HU Berlin zu. Allerdings gibt es auch 21 Repositorien mit insgesamt 149.599 Dokumenten, welche ausschließlich mit dem MIME-Typ *text/html* publizieren.

Im Rahmen dieser Arbeit geht es darum herauszufinden, wie hoch der Anteil an Dokumenten ist, deren Format die direkte Einbindung von OA_SCORE erlaubt. Dies sind vor allem die Formate *application/pdf*, *text/html* und *text/xml*.

Betrachtet man die Verteilung der MIME-Typen in den untersuchten Repositorien in **Abbildung 8** wird bereits deutlich, dass hier dem Format *application/pdf* eine herausragende Bedeutung zukommt. Mit 3.681.128 Nennungen beträgt der Anteil dieses Formats 20,6 %.

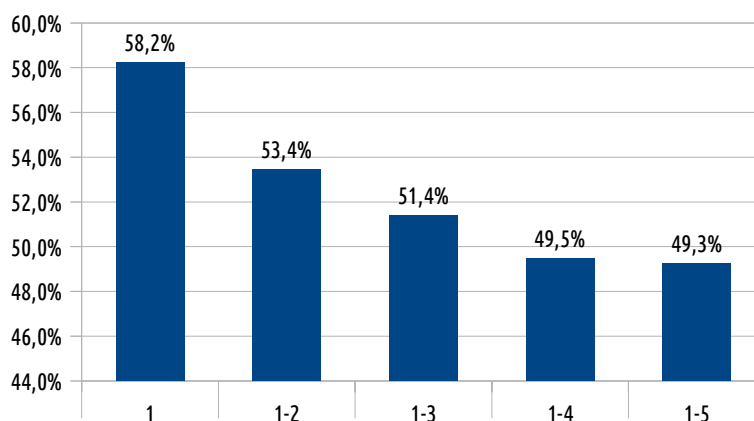


Abbildung 11: Anteil des MIME-Typs *application/pdf* an der Gesamtzahl der Formatangaben nach zusammengefassten Häufigkeitsklassen

Der Anteil wächst auf 58,2 %, wenn nur die Repositorien betrachtet werden, die nach ihrem Format-Record-Verhältnis zur Häufigkeitsklasse 1 gehören. Selbst wenn man alle Repositorien von Häufigkeitsklasse 1 bis Häufigkeitsklasse 3 betrachtet, beträgt der Anteil der Dokumente mit *application/pdf* noch knapp über 50 %. Bei Repositorien der Häufigkeitsklassen 1 bis 3 wird vermutet, dass darin der Anteil aktueller Forschungsveröffentlichungen überdurchschnittlich hoch ist.



Abbildung 10: Verteilung der gefundenen Formatangaben (aufgeführt sind nur Angaben mit $n > 1.000$; In Klammern die Angabe, ob die MIME-Type von IANA anerkannt wird [1] oder nicht [0])

Bezogen auf die Ausgangsfrage kann somit gesagt werden, dass die aktuell vorherrschende Kultur elektronischer Veröffentlichungen mit ihrem starken Anteil an Veröffentlichungen im *Portable Data Format* eine gute Voraussetzung für die Anwendung von OA_SCORE bietet.

Die eingangs erstellte Nullhypothese kann unter der Bedingung widerlegt werden, dass nur Repositorien betrachtet werden, deren Format-Record-Verhältnis konservativ geschätzt unter 3 liegt. Zu dieser Gruppe gehören 944 der untersuchten 1.036 Repositorien, was einem Anteil von 91 % entspricht.

Im weiteren Verlauf wurde gefragt, wie hoch der Anteil der nicht begutachteten Items ist und ob sich daraus Konsequenzen für die Anwendbarkeit von OA_SCORE ergeben.

Angesichts der vorausgegangenen Ergebnisse werden hierzu nur noch die vorliegenden Daten für den MIME-Typ *application/pdf* ausgewertet.

Es wurde entsprechend folgende Verteilung ermittelt:

	<i>N_{application/pdf}</i>
<i>Items, die auf eine Abschlussarbeit hindeuten (T)</i>	812.003
<i>Items, die auf einen Zeitschriftenaufsatz hindeuten (J)</i>	524.488
<i>Sonstige Items (S)</i>	2.361.010
Summe	3.697.501
Anzahl Repositorien	915

Tabelle 10: Verteilung des MIME-Typs *application/pdf* bezogen auf den Typ des Items

Mittels einer Stichprobe wurde nunmehr versucht, das Ergebnis zu validieren. Gemäß der zuvor gefassten Überlegungen sollen nur Repositorien für die Stichprobe zu gelassen werden,

1. bei denen die Formatangabe *application/pdf* gefunden wurde
2. in denen – bezogen auf den MIME-Typ *application/pdf* – mindestens 5 % der gefundenen Formatangaben als „begutachtet“ identifiziert wurden
3. in denen alle zu untersuchenden Records entweder in Englisch, Deutsch oder Französisch abgefasst sind.

Unter Berücksichtigung der Bedingungen 1 und 2 ergeben sich die folgenden Häufigkeiten:

	<i>N_{application/pdf}</i>
<i>Records, die auf eine Abschlussarbeit hindeuten (T)</i>	807.366
<i>Records, die auf einen Zeitschriftenaufsatz hindeuten (J)</i>	522.468
<i>Sonstige Records (S)</i>	1.052.550
Summe	2.382.384
Anzahl Repositorien	546

Tabelle 11: Verteilung des MIME-Typs *application/pdf* bezogen auf den Typ des Records und Repositorien, in denen mindestens 5 % begutachtete MIME-Typen *application/pdf* gezählt wurden.

Für die Stichprobe wurden jeweils die neu erstellten Records der ersten Abfrage von 14 Repositorien (2,5 % von 546 = 13,6) herangezogen. Die Auswahl erfolgt durch die Markierung jedes 40. Repositoriums in der geordneten Liste.

Das Ergebnis der Zählung ist in **Tabelle 12** dargestellt. Wie zu ersehen ist, wurden im Zuge der Kontrollzählung in mindestens drei Repositorien (1271, 1343, 193) Widersprüche gegenüber der automatischen Zählung festgestellt. In allen drei Fällen stimmte das XML-Sche-

ma der Abfrage nicht mit dem XML-Schema überein, das als Grundlage für die automatische Analyse verwendet wurde.

Das Problem besteht – wie nachträglich festgestellt wurde – in der Definition der *descriptive Metadata Section* (dmdSec) des Schemas für *METS*. Das Schema erlaubt hier innerhalb des *xmlData*-Elements die Einbettung verschiedener Formate. Üblicherweise wird dort das *Metadata Object Description Schema* (MODS) eingebettet. Wie jedoch aus der Beschreibung des *xmlData*-Elements hervorgeht, ist die Angabe eines Schemas nicht zwingend vorge-schrieben. In diesem Fall wird der Inhalt nicht validiert, sondern lediglich auf Wohlgeformt-heit überprüft. (**METS 2012, 32**). In den drei genannten Repositorien wurde das Schema für *MODS* in das *xmlData*-Element integriert. Als Konsequenz daraus folgt `<mods:mods>` als direktes Kind-Element. Bei anderen Repositorien (z. B. 1197) besitzt *xmlData* kein Schema-Attribut. Als Kind-Elemente folgen hier die einzelnen Elemente von MODS, u. a. `<mods:genre>`. Bei der Konzeption des BASH-Scripts wurde nur die letztere Gruppe von Schemata berücksichtigt. Dies erfolgte in der Annahme einer einheitlich vorgegeben Be-schreibung für *METS*.

Das BASH-Script müsste also derart vervollständigt werden, dass auch Repositorien der an-deren Gruppe der *METS*-Schemata darin berücksichtigt werden. Im Rahmen dieser Arbeit konnte diese Anpassung und die anschließende Abfrage der Repositorien nicht mehr durch-geführt werden.

ID	URL	N _I	N _{IF}	Repositorium			Stichprobe			N _S	Bemerkungen
				T	J	S	T	J	S		
1197	http://academica-e.unavarra.es/oai/request	2596	3533	2627	0	0	100	0	0	100	
1271	http://brage.bibsys.no/hibu/re-quest	213	430	0	105	111	37	62	1	100	Thesis nicht erkannt, da <code><mods:mods></code> im Path
1343	http://dehesa.unex.es:8080/oai/re-quest	71	139	70	0	5	57	4	0	100	ISSN nicht erkannt, weil Element <code><mods:mods></code> im Path fehlt.
19	http://dl.acs.org.au/index.php/in-dex/oai	472	473	0	473	0	0	100	0	100	
193	http://dyuthi.cusat.ac.in/dspace-oai/request	2273	4699	0	180	2161	105	0	0	100	Thesis nicht erkannt, da <code><mods:mods></code> im Path
803	http://eprints.sdu.edu.tr/cgi/oai2	644	653	653	0	0	98	0	0	100	
1579	http://helvia.uco.es/oai/request	8589	15214	549	0	8317	69	0	23	100	
1972	http://library.wur.nl/oai	1151	1154	35	767	279	4	19	7	50	
1728	http://opus.ub.uni-hohenheim.de/oai2/oai2.php	759	759	456	59	244	50	0	0	100	
1066	http://repositorio.ufsc.br/oai/re-quest	19060	20171	12479	0	4120	0	0	47	100	
491	http://ria.asturias.es/dspace-oai/request	284	292	1	67	196	0	23	59	100	Viele Angaben zu Artikeln konnten nur manuell erkannt werden. Kaum Hinweis auf ISSN
1186	http://sophia.stkate.edu/do/oai/	305	305	185	0	65	0	0	44	100	
736	http://www.biblioteca.ueb.edu.ec/oai/request	951	1703	1123	0	3	97	0	0	100	
894	http://www.opus.ub.uni-erlan-gen.de/opus/oai2/oai2.php	2873	2873	2249	377	247	32	7	11	50	

Tabelle 12: Ergebnis der Kontrollzählung von 14 Repositorien in Bezug auf die Verteilung von begutachteten PDF-Dokumenten

N _I	Anzahl der Records im Repositorium
N _{IF}	Anzahl der Records mit Formatangaben im Repositorium
T	Anzahl der Records vom Typ „Thesis“
J	Anzahl der Records vom Typ „Journal“
S	Anzahl der Records vom Typ „Sonstige“
N _S	Anzahl der Records in der Stichprobe

Es muss nunmehr noch geprüft werden, ob der festgestellte systematische Fehler noch in irgendeiner Weise heraus gerechnet werden kann, um die Nullhypothese zu widerlegen.

Die Lösung für dieses Problem besteht theoretisch darin, nur die Repositorien zu berücksichtigen, in denen der MIME-Typ *application/pdf* sowohl in der Rubrik „Thesis“ als auch in der Rubrik „Journal“ in signifikanter Zahl vorkommt. Dieses Vorgehen berücksichtigt jedoch folgende Sachverhalte nicht:

- Das BASH-Script benutzt noch weitere Kriterien für die Zuordnung zu einer der Rubriken. Es ist auch nach der Kontrollzählung nicht klar, inwieweit diese Kriterien korrekt ausgewertet wurden.
- Mindestens zwei häufig begutachtete Publikationsformen wurden bei der Zählung nicht berücksichtigt. Die betreffenden Records wurden der Kategorie „Sonstige“ zugeordnet. Es handelt sich um Konferenz- und Workshopbeiträge sowie um Forschungsberichte. Beide Publikationsformen werden in der Regel begutachtet oder auch nicht. Eine eindeutige Bestimmung ist anhand der vorliegenden Daten nicht möglich.
- Es gibt Repositorien mit nicht-begutachteten Inhalten. Im Rahmen der Kontrollzählung trifft dies insbesondere auf das Repository 1186 zu.

Damit wird klar, dass die Ausgangsfrage mit der gewählten Methode nicht zufriedenstellend beantwortet werden kann. Die in diesem Zusammenhang aufgestellten Nullhypothesen konnten nicht widerlegt werden. Es ist angesichts der Verschiedenheit der abgefragten Daten zweifelhaft, ob eine detailliertere Analyse eine Lösung darstellt. Neben den systematischen Unterschieden bei der Erzeugung der Ausgaben, müssten hierfür auch individuelle Fehler bei der Eingabe berücksichtigt werden.

Eine Lösung bringt wahrscheinlich nur die Entwicklung und konsequente Anwendung eines Elements, welches über den Status der Begutachtung Auskunft gibt. Dieses Element müsste jedoch konsequent, also sowohl für begutachtete als auch für nicht-begutachtete Items, angegeben werden.

6.2 Bewertungsdesign

6.2.1 Datenkorpus

Der Fragebogen wurde in den Sprachen Deutsch und Französisch entwickelt (**Anhang 10.5**).

Befragt wurden insgesamt 39 Personen, davon 21 direkt und weitere 18 indirekt über das Internet. Letzteres gestaltete sich jeweils als ein längerer Austausch von E-Mail- bzw. Facebook-Nachrichten.

Insgesamt haben 27 Personen die Befragung abgeschlossen. 12 Teilnehmer – allesamt indirekt befragt – haben das Interview abgebrochen. Die hohe Abbruchquote kann zum Teil mit der Komplexität der Fragestellung begründet werden, denn 8 Teilnehmer brachen bei der Beurteilung des ersten Textes ab.

Die Zahl der Verweigerer (*Unit-Nonreponse*) wurde nicht festgehalten.

24 Befragungen wurden mit dem deutschsprachigen Fragebogen durchgeführt, 9 davon wurden abgebrochen. Entsprechend wurden 15 Befragungen mit dem französischsprachigen Fragebogen durchgeführt, wovon wiederum 3 abgebrochen wurden.

Die geringe Zahl der ausgewerteten Fragebögen, schränkt die Aussagefähigkeit der Untersuchung stark ein. Die ausgewerteten Daten sind über den **Anhang 10.10** zugänglich.

6.2.2 Ergebnisse

Bei der Auswertung der Interviewergebnisse sollte zunächst die Frage nach der Vorliebe für bestimmte Begriffe untersucht werden. In **Tabelle 13** ist aufgeführt, wie häufig die einzelnen Begriffe zu Beginn der Befragung genannt wurden.

Was die deutschsprachigen Begriffe betrifft, können jeweils Präferenzen für Plausibilität und Relevanz vermutet werden. Beim Begriffspaar Originalität/Einfallsreichtum ist der Unterschied nicht stark genug ausgeprägt.

Bei den französischen Begriffen werden die Vorlieben für *Originalité* und *Utilité* deutlich, während beim Paar *Crédibilité/Fiabilité* keine Aussage getroffen werden kann.

Vier Befragte änderten die Begriffe im Lauf des Interviews; je zwei in jeder Sprachversion. Bei den beiden Modifikationen in der deutschen Sprachversion lässt sich die Tendenz feststellen, den Begriff Relevanz durch Nützlichkeit zu ersetzen. Außerdem hielten drei Befragte am Ende des Interviews die gewählten Kriterien nicht als Grundlage für eine Bewertung geeignet.

fr		de	
<i>Unicité</i>	0	<i>Originalität</i>	14
<i>Originalité</i>	15	<i>Einfallsreichtum</i>	10
<i>Crédibilité</i>	7	<i>Glaubwürdigkeit</i>	7
<i>Fiabilité</i>	8	<i>Plausibilität</i>	17
<i>Utilité</i>	13	<i>Relevanz</i>	17
<i>Valeur</i>	2	<i>Nützlichkeit</i>	7

Tabelle 13: Verteilung der Vorlieben für bestimmte Kategoriebezeichnungen

6.2.2.1 2D-Skala

Bei der 2D-Skala ging es zunächst um die Definition der Achsen. Relativ deutlich ist zu sehen, dass der Relevanz (R) eine besondere Bedeutung zugeschrieben wird. Plausibilität (P) und Originalität (O) folgen in den Rängen.

	O - P	O - R	P - R	Summe
<i>de</i>	3	7	13	23
<i>fr</i>	2	4	8	14
<i>Summe</i>	5	11	21	37

Tabelle 14: Häufigkeiten der Achsen-Definitionen bei der 2D-Skala (Paare)

	O	P	R	Summe
<i>de</i>	10	16	20	46
<i>fr</i>	6	10	12	20
<i>Summe</i>	16	26	32	66

Tabelle 15: Häufigkeiten der Achsen-Definitionen bei der 2D-Skala (Einzelne Kriterien)

Es wurden nun die Anteile der jeweiligen Kategorien an den Aussagen extrahiert und dar-

aus das arithmetische Mittel abgeleitet. Das heißt z. B. bei einem Wert x, y in einem Diagramm mit den Achsen Plausibilität (y) und Relevanz (x), dass die Werte y für Plausibilität und x für Relevanz extrahiert wurden. Dabei können die Werte zwischen 0 (Minimum) und 3 (Maximum) liegen.

Anschließend wurde nochmals ein globales arithmetisches Mittel aus den Mittelwerten aller Texte gebildet, um die Texte miteinander vergleichen zu können.

Bei einer rein zufälligen Verteilung sowie $n \rightarrow \infty$, müssten alle arithmetischen Mittel den Wert 1,50 annehmen. Aus **Tabelle 16** wird deutlich, dass dem nicht so ist.

	<i>n</i>	<i>Text1</i>	<i>Text2</i>	<i>Text3</i>	<i>Text4</i>	<i>Text5</i>	$\bar{x}_{arithm} (Text1:Text5)$
Justierung		O=0,00	P=0,00	R=0,00	O=P=R=0,00	O=P=R=3,00	
Originalität	9 - 10	0,91	1,00	1,30	0,90	1,60	1,14
Plausibilität	21 - 22	1,86	0,36	1,77	1,18	2,05	1,45
Relevanz	23 - 25	1,64	0,83	0,50	0,54	2,17	1,14
\bar{x}_{arithm}		1,47	0,73	1,19	0,87	1,94	

Tabelle 16: Arithmetische Mittel der einzelnen Kategorien aus der 2D-Skala für die untersuchten Texte

Alle angenommen Ergebnisse sind – mit Einschränkungen – tatsächlich eingetreten. Bei den Texten 1 bis 3 erhielten die jeweils als falsch gesetzten Kategorien den kleinsten Mittelwert. Bei Text 4 liegen alle Mittelwerte unter dem globalen Mittel, bei Text 5 liegen alle Mittelwerte darüber.

Jedoch wird auch deutlich, dass kein Wert innerhalb des Bereiches von 5 % um den erwarteten Wert herum liegt. Im Gegenteil, der Wert für Originalität beim Text 5 liegt noch im Bereich von 5 % um den rein zufälligen Erwartungswert. Angesichts der geringen Zahl an Interviews dürfte der Fehlerbereich zudem mehr als 5 % betragen. Für die Abweichungen kommen zwei mögliche Ursachen in Frage:

- Die Justierung der Texte war nicht ausreichend. So kann z. B. auch einem Nonsens-Text noch Information entnommen werden, solange diese in Form von als Wörtern zu erkennenden Zeichenfolgen vorliegt. Im Fall von Text 2 kann der vom Ideal abweichende Wert für Plausibilität, entweder damit zusammenhängen, dass ein Teil der Befragten die Ansichten Hegels teilt (sehr unwahrscheinlich) oder dass die Bewertung auf den historischen Kontext des 19. Jahrhunderts bezogen wurde.
- Die Skala wurde von einem Teil der Befragten schlichtweg nicht verstanden. Für diese Vermutung steht die Beobachtung, dass mehrere Befragte keinen naturwissenschaftlich-mathematischen Hintergrund haben und Schwierigkeiten beim Verständnis der Skala zur Sprache brachten.

Einige weitere Auffälligkeiten müssen diskutiert werden:

- Der Wert für die Originalität zum Text 1 liegt mit 0,91 näher beim zufälligen als beim erwartetem Wert. Dies deutet darauf hin, dass der Begriff der Originalität von den Befragten sehr unterschiedlich aufgefasst wurde. Es kann aber auch sein, dass die Bewertungen des ersten Texts grundsätzlich fehlerbehafteter sind, weil die Befragten sich erst an die Form der Bewertung gewöhnen mussten. Für die erste Vermutung steht auch der Umstand, dass im Text 5 der Wert für die Originalität ebenfalls nahe am zufälligen Erwartungswert liegt.
- Die Plausibilität von Text 3 wurde durchschnittlich nur mit 1,77 angegeben. Obwohl zwar mit diesem Text nur die Nicht-Relevanz getestet werden sollte, war der Text an sich sehr plausibel. Zwei Ursachen können vermutet werden. Erstens der Umstand, dass der Text als nicht relevant eingeschätzt wurde. Dadurch waren die Befragten auch nicht ausreichend motiviert, die Plausibilität zu bewerten. Zweitens die Wahl

der Quellenangabe *www.wikipedia.org*. Einige Befragte erwähnten, dass sie die Online-Enzyklopädie grundsätzlich nicht als zuverlässige Quellenangabe betrachten.

- Die ermittelten Werte für Text 4 liegen allesamt zu hoch. Dies trifft insbesondere für den Wert der Plausibilität zu. Das kann damit zusammenhängen, dass der Text viele Begriffe aus der Informationstechnik verwendet und Befragte ohne grundlegende Kenntnisse auf diesem Gebiet allein die Nennung von Fachbegriffen als Kriterium für eine hohe Plausibilität erachteten.

Es ist wichtig festzustellen, dass der Nonsens-Charakter von Text 4 nur zu ermitteln ist, wenn ein globales Mittel mit der Bewertung aller Texte zur Verfügung steht. Auf keinen Fall kann der zufällige Mittelwert von 1,50 als Maßstab dienen, da dieser nur theoretisch hergeleitet ist und das persönliche Bewertungsverhalten von Mitgliedern einer bestimmten Gemeinschaft nicht widerspiegelt.

Nach der bisherigen Darstellung ließe sich die 2D-Skala mit den folgenden wesentlichen Einschränkungen für OA_SCORE nutzen:

- Die ermittelten Werte können nicht vom Verdacht befreit werden, stark fehlerbehaftet zu sein.
- Es bedarf einer ausreichend großen Anzahl von Bewertungen, um die Fehlerwahrscheinlichkeit zu reduzieren. In jedem Fall muss die Anzahl der Bewertungen deutlich erwähnt werden, damit die Nutzer die getroffenen Aussagen einordnen können.
- Die angegebene Genauigkeit der ermittelten Werte muss die Fehlerwahrscheinlichkeit berücksichtigen. Die Angabe von zwei Nachkommastellen ist z. B. bei weniger als 100 Bewertungen nicht gerechtfertigt.

Um die genannten Nachteile auszugleichen, besteht die Möglichkeit, nicht das arithmetische Mittel, sondern den Median zu berechnen. Dieses Mittel ist robuster gegenüber Ausreißern und kann auch bei einer geringen Anzahl an Bewertungen noch plausible Aussagen liefern. Die Ergebnisse sind in **Tabelle 17** dargestellt.

	n	Text1	Text2	Text3	Text4	Text5
Justierung		O = 0	P = 0	R = 0	O = P = R = 0	O = P = R = 3
Originalität	9 - 10	1	0	1	0	2
Plausibilität	21 - 22	2	0	2	1	2
Relevanz	23 - 25	2	0	0	0	2

Tabelle 17: Median der einzelnen Kategorien aus der 2D-Skala für die untersuchten Texte (bei geraden N wurde der Untermedian verwendet)

Die Abweichungen von den erwarteten Werten sind immer noch sichtbar ($O_{\text{Text 1}}$, $P_{\text{Text 4}}$). Da jedoch die Information reduziert wurde, erscheinen die Aussagen angesichts der geringen Anzahl von Bewertungen plausibler. Dies sei hier abschließend durch den Vergleich mit dem auf ganze Zahlen gerundeten arithmetischen Mittelwerten veranschaulicht. Insbesondere beim Text 4 kommt der Median dem erwarteten Ergebnis näher als das arithmetische Mittel.

	n	Text1	Text2	Text3	Text4	Text5
Justierung		O = 0	P = 0	R = 0	O = P = R = 0	O = P = R = 3
Originalität	9 - 10	1	1	1	1	2
Plausibilität	21 - 22	2	0	2	1	2
Relevanz	23 - 25	2	1	0	1	2

Tabelle 18: Arithmetisches Mittel der einzelnen Kategorien aus der 2D-Skala für die untersuchten Texte, jeweils auf ganze Zahlen gerundet (vergl. Tabelle 17).

Eine weitere Möglichkeit der Annäherung wären getrimmte arithmetische Mittel. Diese könnten z. B. durch Weglassen von jeweils 5 % der Werte an beiden Extremen ermittelt

werden. Für Text 4 ergäbe sich $O=0,75$, $P=1,15$ und $R=0,45$ bzw. gerundet $O=1$, $P=1$ und $R=0$.

Im Rahmen dieser Arbeit kann jedoch keine abschließende Untersuchung über das anzuwendende Berechnungsverfahren durchgeführt werden.

6.2.2.2 3D-Skala

Die 3D-Skala lässt zwei Arten von Aussagen zu. Einerseits wird die Kategorie hervorgehoben, die dem Befragten besonders zutreffend erscheint. Andererseits lässt sich eine Aussage über den Abstand dieser ausgewählten Kategorie zu den anderen beiden Kategorien ermitteln. Dazu kann das Teilfeld einer Kategorie (hier: grün, blau oder rot) als eine 2D-Skala mit den Werten 1,2 bzw. 3 auf den beiden Achsen aufgefasst werden. Die Bezeichnung der Achsen ergibt sich aus der auf dieser Seite benachbarten Kategorie. In **Abbildung 12** wird das Vorgehen erläutert.

Was die Analyse der bevorzugten Kategorie betrifft, wurde jede Nennung mit dem Wert 1 belegt, jede Nichtnennung entsprechend mit dem Wert 0. Es wurde nicht berücksichtigt, welches Feld innerhalb des Teilbereichs ausgewählt wurde. Anschließend wurden wiederum die arithmetischen Mittel für die einzelnen Texte sowie ein globales Mittel für alle Texte berechnet. Das Ergebnis ist in **Tabelle 19** dargestellt. Es muss beachtet werden, dass die Auswahl-Wahrscheinlichkeit für jede Kategorie nicht $\frac{1}{2}$, sondern nur $\frac{1}{3}$ beträgt.



Abbildung 12: Erläuterung zur Analyse der Werte aus der 3D-Skala.

Im Beispiel wurde das Feld 3 ausgewählt. Es liegt im Teilbereich „Originalität“, womit $O=1$ und $P=R=0$ gesetzt werden. Hierbei wird nicht die genaue Lage dieses Feldes berücksichtigt.

Stellt man sich hingegen den Teilbereich „Originalität“ als ein 2D-Diagramm mit den Achsen Plausibilität (x) und Relevanz (y) vor, so hätte die Aussage zum Feld 3 den Wert (3,2). Der Abstand zum Teilbereich „Plausibilität“ beträgt 3 Felder, der zum Teilbereich „Relevanz“ 2 Felder. Entsprechend können die gerichteten Abstandswerte $O \rightarrow P=3$ und $O \rightarrow R=2$ extrahiert werden. (der Pfeil ist als „Abstand von ... nach ...“ zu lesen)

	n	Text1	Text2	Text3	Text4	Text5	$\bar{x}_{arithm} (Text1:Text5)$
Justierung		O = 1,00	P= 1,00	R=1,00	O=P=R=1,00	O=P=R = 4,00	
Originalität	27 - 29	0,31	0,46	0,32	0,43	0,19	0,34
Plausibilität	27 - 29	0,38	0,14	0,54	0,43	0,33	0,36
Relevanz	27 - 29	0,31	0,39	0,14	0,14	0,48	0,29
\bar{x}_{arithm}		0,33	0,33	0,33	0,33	0,33	

Tabelle 19: Arithmetische Mittel der einzelnen Kategorien aus der 3D-Skala für die untersuchten Texte

Wie zu erkennen ist, beinhalten die ermittelten Werte größere Fehler, als dies für die 2D-Skala zutrifft. Die Bewertung für die Originalität des ersten Textes liegt z. B. zu nah am zufälligen Mittelwert, so dass sie nicht berücksichtigt werden kann.

Bei den Texten 4 und 5 wird deutlich, dass mit dieser Methode keine absolute Aussage bezüglich aller drei Kriterien möglich ist. Es könnte hier höchstens ermittelt werden, wie nah

die Bewertungen für die einzelnen Kriterien beieinander liegen. Im Idealfall würde hier jeweils 0,33 stehen, ganz gleich, ob es sich um einen Nonsens-Text handelt oder nicht.

Im Folgenden wurden auch für die gerichteten extrahierten Abstandswerte (wie in **Abbildung 12** beschrieben) die arithmetischen Mittel gebildet. Genau genommen wäre die Berechnung eines arithmetischen Mittelwertes hier gar nicht zulässig, da es sich um eine Ordinalskala handelt.

	$O \rightarrow P$	$O \rightarrow R$	$P \rightarrow O$	$P \rightarrow R$	$R \rightarrow O$	$R \rightarrow P$
Text1	2,11	2,00	2,27	1,91	2,22	2,33
Text2	2,54	2,62	3,00	2,00	2,09	1,82
Text3	1,89	1,78	2,07	2,40	1,75	2,00
Text4	2,25	2,25	2,00	2,00	3,00	2,00
Text5	2,20	1,80	1,67	1,89	1,38	1,85

Tabelle 20: Arithmetische Mittel der gerichteten extrahierten Abstandswerte

Der Wertebereich liegt hier zwischen 1,00 (sehr nah) und 3,00 (sehr weit). Bei einer zufälligen Verteilung beträgt der erwartete arithmetische Mittelwert 2,00.

In einem nächsten Schritt wurden die beiden Paare einer Beziehung durch Subtraktion zusammengefasst. Negative Werte stellen einen Abstand zugunsten des rechten Kriteriums dar, positive Werte einen Abstand zugunsten des linken Kriteriums. Der mögliche Wertebereich liegt zwischen -2,00 und +2,00. Bei einer zufälligen Verteilung beträgt die Differenz folglich 0,00. Zusätzlich wurde das arithmetische Mittel der drei Beziehungen pro Text gebildet.

	$O \rightarrow P$	$O \rightarrow R$	$R \rightarrow P$	\bar{x}_{arithm}
Text1	-0,16	-0,22	+0,42	+0,01
Text2	-0,46	+0,52	-0,18	-0,04
Text3	-0,18	+0,03	-0,40	-0,18
Text4	+0,25	-0,75	+0,00	-0,17
Text5	+0,53	+0,42	-0,04	+0,30

Tabelle 21: Zusammengefasste arithmetische Mittel der gerichteten extrahierten Abstandswerte

Wie bereits zuvor, als die Häufigkeit von besonders zutreffenden Kriterien untersucht wurde, lassen sich auch hier kaum belastbare Aussagen treffen. Zwar liegen bei den Texten 1 bis 3 die jeweils „falsch“ gesetzten Kriterien am weitesten von den beiden anderen Kriterien entfernt, bei den Texten 4 und 5 versagt jedoch dieses Verfahren schlichtweg seinen Dienst. Die Skala kann Nonsens-Texte nicht von „gültigen“ Texten unterscheiden.

Zusammenfassend kann gesagt werden, dass die 3D-Skala in dieser Form nicht zur Kurzbewertung von wissenschaftlichen Texten geeignet ist.

6.2.2.3 1-von-8-Skala

Zur Analyse der 1-von-8-Skala wurden ebenfalls zunächst die Werte für jede Kategorie extrahiert. Dabei bedeutet der Wert 0 „trifft nicht zu“ und der Wert 1 „trifft zu“. Die arithmetischen Mittel der Aussagen sind in **Tabelle 22** dargestellt.

	Text1	Text2	Text3	Text4	Text5	\bar{x}_{arithm}
Justierung	$O=0,00$	$P=0,00$	$R=0,00$	$O=P=R=0,00$	$O=P=R=1,00$	
Originalität	0,45	0,25	0,25	0,50	0,52	0,39
Plausibilität	0,59	0,14	0,79	0,43	0,96	0,58
Relevanz	0,72	0,25	0,36	0,14	0,67	0,43
\bar{x}_{arithm}	0,59	0,21	0,46	0,36	0,72	

Tabelle 22: Arithmetische Mittel der aus der 1-von-8-Skala extrahierten Werte (n: 27 - 29)

Da die Aussagen nur zwei Merkmalsausprägungen zulassen, entspricht der Untermedian dem auf die ganze Zahl gerundeten arithmetischen Mittel.

	Text1	Text2	Text3	Text4	Text5
	<i>O=0</i>	<i>P=0</i>	<i>R=0</i>	<i>O=P=R=0</i>	<i>O=P=R=1</i>
<i>Originalität</i>	0	0	0	0	1
<i>Plausibilität</i>	1	0	1	0	1
<i>Relevanz</i>	1	0	0	0	1

Tabelle 23: Untermedian der aus der 1-von-8-Skala extrahierten Werte (n: 27 - 29)

Nach **Tabelle 23** treffen alle erwarteten Werte ein. Jedoch liegen – wie aus **Tabelle 22** ersichtlich ist, bei den Texten 1, 4 und 5 die Werte für Originalität im Bereich von 5 % um den erwarteten Wert bei einer zufälligen Verteilung. Da das globale arithmetische Mittel für die Originalität aller Texte auffallend niedrig ist, dürfte die Ursache hierfür bei der Definition bzw. dem Verständnis des Kriteriums liegen. Dass die Originalität des ersten Textes überdurchschnittlich hoch bewertet wird, kann mit einer Eingewöhnungsphase erklärt werden. In der Tat waren nach eigenen Aussagen mehrere Befragte durch den Aufbau der Skala in zwei untereinander liegende Blöcke zu je vier Bewertungsmustern irritiert. Die horizontale Trennlinie zwischen den beiden Blöcken wurde irrtümlich als die Linie für den Bewertungsmaßstab angesehen.

Nach den bisherigen Darstellungen ließe sich die 1-von-8-Skala mit folgenden Einschränkungen bzw. Verbesserungen für OA_SCORE nutzen:

- Die Anzahl der Bewertungen muss genannt werden.
- Die verwendeten Genauigkeiten müssen in einem sinnvollen Zusammenhang mit der Anzahl der Bewertungen stehen.
- Die acht Bewertungsmuster sollten nebeneinander dargestellt und deutlich voneinander getrennt werden.

6.2.3 Zusammenfassung zum Bewertungsdesign

Die Befragten wurden am Ende des Interviews gebeten, die von ihnen bevorzugte Skala zu nennen. Die Ergebnisse sind in **Tabelle 24** dargestellt.

Skalentyp	Nennungen
<i>2D-Skala</i>	9
<i>3D-Skala</i>	8
<i>1-von-8-Skala</i>	10

Tabelle 24: Von den Befragten bevorzugte Skalen (n=27)

Die gegebenen Antworten lassen keine belastbare Aussage zu, zumal die 3D-Skala ohnehin nicht verwendbar ist. Es kann vermutet werden, dass bei einigen Antworten ästhetische Gründe ausschlaggebend waren.

Aus den bisherigen Beobachtungen lässt sich jedoch eine Entscheidungshilfe zwischen 2D-Skala und 1-von-8-Skala ableiten.

	2D-Skala	1-von-8-Skala
<i>Zahl der Kriterien, die gleichzeitig abgefragt werden können</i>	2	3
<i>Genauigkeit bei geringer Zahl von Bewertungen</i>	<i>Ausreichend, wenn ganzzahlige Werte verwendet werden</i>	<i>Ausreichend, wenn ganzzahlige Werte verwendet werden</i>
<i>Wertebereich der Aussage</i>	4 (0,1,2,3)	2 (0,1)
<i>Besonderheiten</i>	<i>Das zu Grunde liegende Prinzip des kartesischen Koordinatensystems ist für Benutzer außerhalb des STM-Bereichs u.U. nicht nachvollziehbar.</i>	<i>Das zu Grunde liegende Prinzip der Kombination von Merkmalsausprägungen ist gewöhnungsbedürftig.</i>

Tabelle 25: Vergleich zwischen 2D-Skala und 1-von-8-Skala

Insbesondere aufgrund der Tatsache, dass mit der 1-von-8-Skala drei Kriterien zugleich abgefragt werden können, erscheint diese von den drei untersuchten Skalen für OA_SCORE die Geeignetste zu sein.

Unabhängig davon, bedarf es einer genaueren Untersuchung hinsichtlich der abzufragenden Kategorien. Die in der Untersuchung verwendeten Kategorien Originalität, Plausibilität und Relevanz scheinen nicht geeignet bzw. sind erläuterungsbedürftig. Anstelle der Originalität scheint ein Rückgriff auf den Begriff „Neuigkeitswert“ angebracht, auch wenn die Synonymanalyse dies nicht zu belegen scheint. Zumindest im Alltagsverständnis des Deutschen beinhaltet Originalität häufig auch das Konzept von „sich amüsieren“. Relevanz benötigt einen objektiven Bezugsrahmen. Dieser könnte im Rahmen von OA_SCORE über eine Auswertung der semantischen Beziehung zwischen der zu bewertenden Arbeit und den seitens des Bewertenden bereits veröffentlichten Arbeiten geschaffen werden. Auch das Konzept der Plausibilität lässt sich nur anwenden, wenn auf Seiten der Bewertenden ein gewisses Vorwissen vorausgesetzt werden kann.

Von den beiden in diesem Zusammenhang aufgestellten Nullhypothesen lässt sich die erste durch die vorliegende Untersuchung nicht widerlegen, während die zweite mit Einschränkungen widerlegt werden kann. **OA_SCORE bietet grundsätzlich die Möglichkeit plausible Aussagen zu erhalten und kann zur Kurzbewertung von wissenschaftlichen Texten eingesetzt werden.** Die Einschränkungen sind:

- Die Aussagekraft von OA_SCORE erlaubt keine exklusive Verwendung. Die Aussagen müssen mit den Aussagen anderer Instrumente und Filter verglichen werden.
- Es muss stets die Anzahl der Bewertungen angegeben werden.

Einige zuvor angestellten Überlegungen wurden bei der Auswertung nicht mehr berücksichtigt, da sie nicht mehr wesentlich zum Erkenntnisgewinn betragen. Dabei handelt es sich insbesondere um die Daten aus Frage 2 des Fragebogens, sowie die Daten der Zeitmessung. Beide Datenreihen sind jedoch im **Anhang 10.10** dargestellt.

7 Weitere Überlegungen zum Design

7.1 Darstellung der Skala

Damit die 1-von-8-Skala zu aussagekräftige Bewertungen führt, sollten folgende Punkte berücksichtigt werden.

Die Skala sollte als **erste Seite** in PDF-Dokumenten eingefügt werden. So kann gewährleistet werden, dass das Bewertungsangebot von allen Lesern wahrgenommen wird.

Ein **Anleitungstext** muss zwingend in unmittelbarer Nähe zur Skala angezeigt werden. Auf keinen Fall sollte dieser Text nur über einen Link zu erreichen sein. So kann die Wahrscheinlichkeit erhöht werden, dass die Benutzer die Skala korrekt anwenden.

Die Darstellung der Skala sollte farbig erfolgen. Die **Farben** müssen auch als Grauwerte kontrastreich dargestellt werden können, damit sie z. B. auf elektronischem Papier bzw. auch von Personen mit Farbfehlsichtigkeit erkannt werden können.

Die Begriffe für die Bewertungskriterien sollten neben Englisch auch in den **Sprachversionen** angeboten werden, die von den wichtigsten Zielgruppen des Repositoriums nachgefragt werden. Eine Kopplung mit der Sprachversion des Dokuments zuzüglich Englisch als Alternative bietet sich an.

Die auszuwählenden Felder sollten in einer einzigen **Reihe** angeordnet werden. Die Legende sollte sowohl am linken als auch am rechten Rand positioniert werden. Dadurch kann sie auch von Benutzern mit Sozialisation im Rechts-Links-Schriftsystem am Beginn der Reihe gefunden werden.

Anstatt mit einer Ziffer sollten die einzelnen Auswahlmöglichkeiten zusätzlich mit **Buchstaben** gekennzeichnet werden, um Verwechslungen mit einer Ordinal- bzw. Intervallskala auszuschließen.

Abbildung 13 zeigt ein Beispiel für eine verbesserte 1-von-8-Skala.

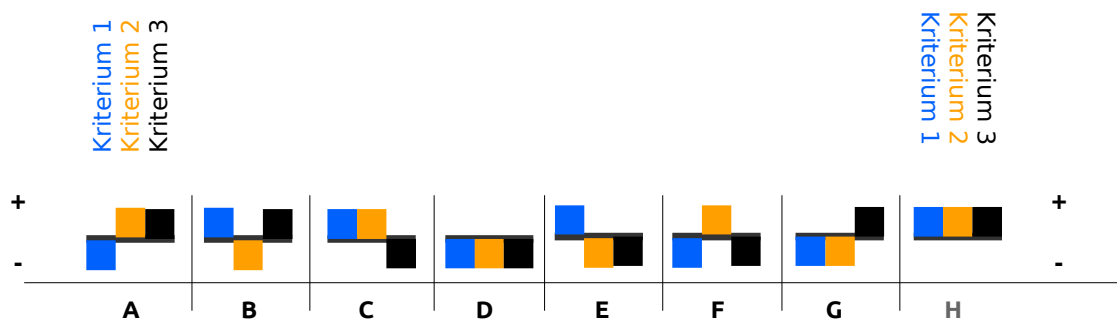


Abbildung 13: Beispiel einer verbesserten verweissensitiven 1-von-8-Skala (ohne Anleitungstext)

7.2 Implementierung

Die Frage nach der Implementierung von OA_SCORE in Repositorien umfasst mehrere Aspekte.

In **technischer** Hinsicht muss die Repositorium-Software so ausgelegt sein, dass sie die Installation von Plug-ins ermöglicht. Die beiden – laut *OpenDOAR* – häufigsten Anwendungen *DSpace* und *EPrints* verfügen über diese Möglichkeit. Bei *DigitalCommons* und *Greenstone*, die mit Abstand folgen, ist eine Implementierung nur durch die Entwickler möglich.

Die Implementierung in das Repositorium eines Data-Providers muss an mindestens vier Stellen erfolgen:

- Bei der Bereitstellung der Skala, entweder zum Zeitpunkt des Downloads bzw. zum Zeitpunkt der Anzeige von Dokument oder Metadaten am Bildschirm.

- Bei der Datenübernahme der Abfrageergebnisse vom Service-Provider. Alternativ können die vom Service-Provider bereitgestellten Daten mittels *Inlineframe* auf der Metadaten-Seite des Data-Providers eingebettet werden.
- Bei der Darstellung des Zwischenergebnisses auf der Übersichtsseite zum Item, falls keine Einbettung über *Inlineframe* erfolgt
- In den allgemeinen Programmoptionen bzw. bei der Interaktion zur Bereitstellung von Items

Zusätzlich sollte eine Möglichkeit angeboten werden, dass Autoren oder andere interessierte Gruppen externe Links zu ihren Veröffentlichungen zusätzlich mit OA_SCORE anreichern.

Es müssen zudem **rechtliche Fragen** geklärt und umgesetzt werden. Dies betrifft insbesondere die Zustimmung bzw. Ablehnung einer Bewertung durch den Autor bzw. weiterer beteiligter Personen (z. B. Prüfer, Gutachter, usw.). Eine Mindestvoraussetzung ist demnach, dass für jedes einzelne Item festgelegt werden kann, ob eine Bewertung mittels OA_SCORE erfolgen soll oder nicht.

Ein weiteres Problem betrifft die Verwendung der gewonnenen Daten. Dürfen negative Bewertungen als Entscheidungsgrundlage bei der Vergabe von Fördergeldern benutzt werden? Soll es einem Autor bzw. einer beteiligten Person erlaubt sein, den Bewertungsprozess nachträglich zu stoppen, wenn das Urteil der Leser ungünstig ausfällt? Wer verfügt überhaupt über die erhobenen Daten und unter welcher Lizenz macht er sie anderen zugänglich? Diese und weitere Fragen können im Rahmen dieser Arbeit nur als Denkanstoß dienen, eine schlüssige Antwort sollte hier nicht erwartet werden. Jedoch ist m. E. eine Bewertung überhaupt nur dann sinnvoll, wenn alle Ergebnisse – auch die negativen – uneingeschränkt der Öffentlichkeit zur Verfügung gestellt werden.

Damit einher geht auch das Problem des **Schutzes vor Manipulationen**. Insbesondere müssen massive Mehrfachbewertungen verhindert werden können. Dies ist letzten Endes nur durch eine Identifizierung der Bewertenden möglich. Eine Nutzeridentifizierung über Konten ist vom Aufwand her vertretbar, wenn die Auswertung über einen zentralen Service-Provider erfolgt. Der Aufwand bei der Registrierung kann durch die Verwendung von OpenID und Browser-Cookies reduziert werden. Durch zusätzliche Verwendung der ORCID können zudem Aussagen zur Aussagefähigkeit der Bewertenden gewonnen werden.

Wichtigste Aufgabe der Nutzeridentifizierung wäre es, mehrfache Bewertungen durch den gleichen Nutzer zu verhindern. Da eine Person mehrere OpenID benutzen kann, lassen sich Mehrfachbewertungen zwar grundsätzlich nicht verhindern, jedoch in der Durchführung erschweren. Einen besseren Schutz bietet hier die ausschließliche Verwendung der ORCID an.

Hinweise auf Manipulationen ergeben sich, wenn die Zeitspanne zwischen Download des Dokuments und dessen Bewertung zu kurz ist, wenn eine ungewöhnliche zeitliche Häufung von Bewertungen festgestellt werden kann oder wenn das Verhältnis zwischen der Zahl der Bewertungen und der Zahl der Downloads außergewöhnlich hoch ist.

In Bezug auf die **Datenintegralität** muss berücksichtigt werden, dass sich die Hashwerte bei jeder mit OA_SCORE angereicherten Kopie ändern. Die Lösung besteht darin, immer auch das unveränderte Originaldokument zum Download anzubieten.

Die zusammengefasste Bewertung der Items muss in geeigneter Form den Benutzern **visualisiert** werden. Ziel sollte es dabei sein, mit möglichst wenigen Aussagen dem Leser derart ausreichend über die Qualität des Items zu informieren, dass dieser dadurch in seiner Entscheidung, sich mit diesem Item näher zu beschäftigen, unterstützt wird. Die unbedingt zu übermittelnden Daten sind die aktuellen Werte für die einzelnen Kriterien sowie die Anzahl der bisherigen Bewertungen. In Erweiterung ist es denkbar, dass die zeitliche Entwicklung

der Bewertungen in Form eines Diagramms eingeblendet wird. Dadurch können zusätzliche Informationen, z. B. über die Rezeptionsgeschichte eines Items übermittelt werden.

Zusätzlich ließe sich – falls die Identifizierung der Benutzer über ORCID erfolgt – die Bewertung auf Kollegen aus dem gleichen Fachbereich einschränken. Hierzu müssten die über ORCID zugänglichen Informationen zu den Bewertenden in geeigneter Weise ausgewertet werden.

Die Bewertung sollte per Item erfolgen, auch wenn dieses mehrere Dokumente beinhaltet.

Es sollte grundsätzlich angestrebt werden, dass die wichtigsten per OA_SCORE gewonnen Daten maschinell von Dritten ausgelesen werden können. Hierbei handelt es sich um die – bei Bedarf - anonymisierten Einzelbewertungen sowie um die zusammengefassten Bewertungen. **Anhang 10.11** enthält ein Datenmodell, ein XML-Schema sowie ein Abfrage-Beispiel hierzu.

Das maschinelle Auslesen sollte entweder für das gesamte Repositorium oder für eine bestimmte Gruppe von Datensätzen ermöglicht werden. Als gemeinsames Merkmal der Gruppe wäre u. a. der Name eines Autors, einer Forschungseinrichtung oder einer Zeitschrift denkbar.

Zusätzlich sollten die bereitzustellenden Daten über noch zu definierende Ontologien mit anderen Ressourcen verknüpft werden können.

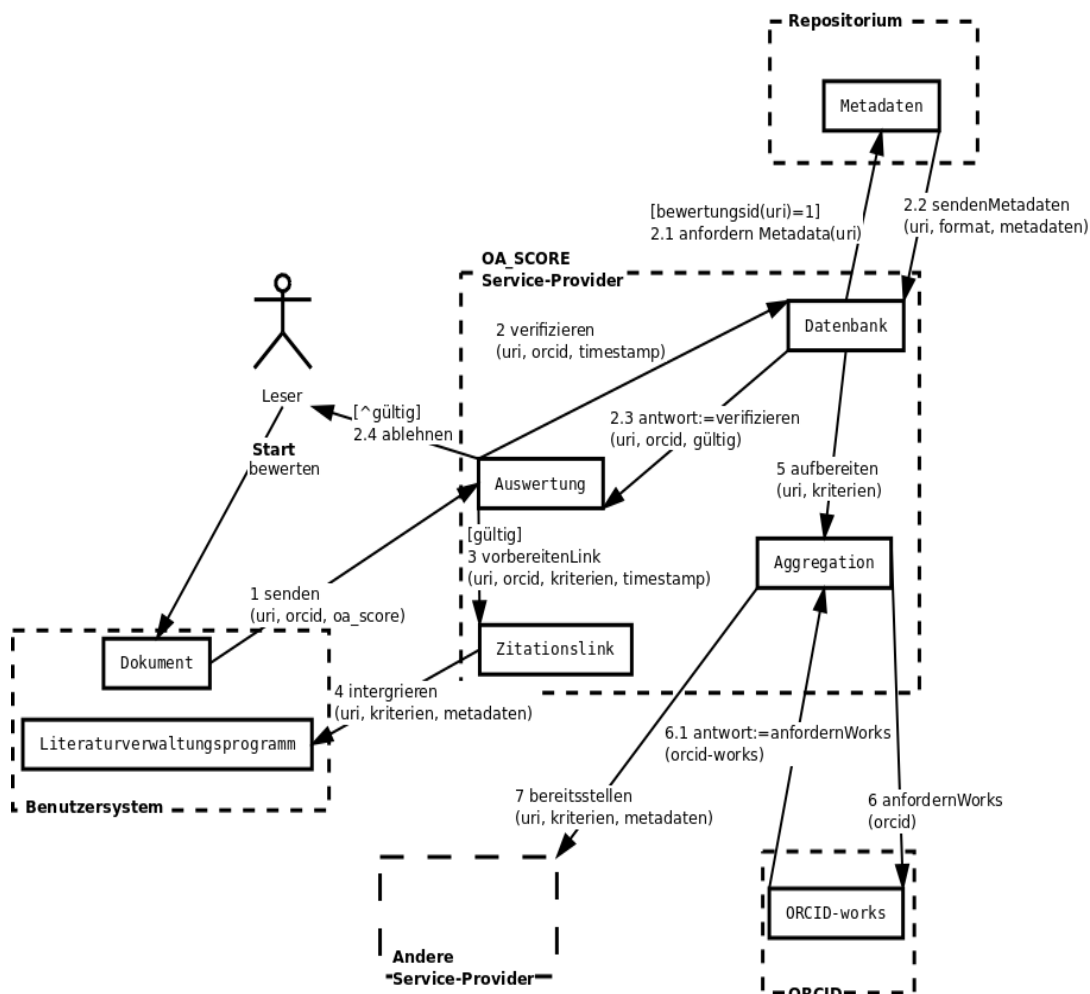


Abbildung 14: Kollaborationsdiagramm des OA_SCORE Service Provider

7.3 Motivation der Teilnehmer

Bereits zu Beginn dieser Arbeit wurde erwähnt, dass der Erfolg alternativer Bewertungsformen von der Bereitschaft der Leser abhängt, daran mitzuwirken. Daraus leitet sich die Frage ab, wie die Benutzer zur Bewertung motiviert werden können, obwohl sie selbst zu diesem Zeitpunkt keinerlei Vorteil dadurch erfahren. Eine Art Einsicht in die gesellschaftliche Notwendigkeit kann man beim Leser nicht voraussetzen.

Ein erster, hier bereits dargestellter Ansatz ist es, die Skala unmittelbar dort zu platzieren, wo die Bewertung stattfinden soll, nämlich im zu bewertenden Dokument selbst. Damit soll der Aufwand der Bewertung reduziert werden. Der Leser muss nach der Lektüre eines Artikels nicht erst zu einer Bewertungsseite zurückgeführt werden.

Dass für die Abgabe der Bewertung eine Identifizierung notwendig ist, stellt eine wichtige Hürde dar. Mit Hilfe von Cookies und OpenID lässt sich das Hindernis etwas abschwächen. Cookies können dafür sorgen, dass sich der Bewertende nicht jedes Mal neu anmelden muss. Eine OpenID verringert den Aufwand der Registrierung. Wird statt der OpenID ausschließlich die ORCID zur Identifizierung eingesetzt, lässt sich wegen des Mehrwerts der Anreiz für eine Neuregistrierung erhöhen.

Die folgenden Überlegungen geben in knapper Form Ideen wieder, wie diese Motivation noch weiter gesteigert werden kann. Sie gehen weit über den bisherigen Ansatz hinaus.

Der Grundgedanke hierbei ist, dass der Leser die abgegebene Bewertung zugleich auch für sich selbst nutzen kann. Ein solcher Fall ist die Organisation der Daten im eigenen Literaturverwaltungsprogramm. Dazu folgende kurze Erläuterung (**Abbildung 14**).

Der Leser lädt zunächst das Dokument wie gewohnt aus dem Repositorium herunter. Nach der Lektüre gibt er seine Bewertung ab. Die Bewertung wird per *HTTP-GET* zum Service-Provider gesendet. Der Server antwortet mit einer Bestätigungsseite, die den Benutzer zunächst über den erfolgreichen Abschluss der Bewertung informiert. Zusätzlich bietet diese Seite - auf die jeweiligen Literaturverwaltungsprogramme angepasste - Zitationslinks an. Der Leser hat nun die Möglichkeit die Metadaten des Items zuzüglich der gerade durchgeführten Bewertung in seine bibliographische Datenbank zu laden. Voraussetzung ist natürlich, dass die jeweiligen Literaturverwaltungsprogramme einerseits Felder für die individuelle Bewertung vorsehen, andererseits aber auch anbieten, nach diesen Bewertungen zu sortieren. Alle gängigen Programme bieten hier z. B. die Möglichkeit, Stichwörter anzulegen und zu verwalten. Im Falle der 1-von-8-Skala genügen hierfür sechs verschiedene Stichwörter, nämlich „Kriterium 1 positiv“; „Kriterium 1 negativ“, usw.

Bereits zuvor, bzw. anlässlich der ersten Bewertung eines Items, werden durch den OA_SCORE-Service-Provider die Metadaten importiert. Basierend auf diesen Daten kann zusätzlich eine spezielle Suchmaschine für bewertete OA-Veröffentlichungen angeboten werden.

Aus Sicht der Anbieter von Literaturverwaltungsprogrammen ist das beschriebene Verfahren zunächst wenig einleuchtend. Interessanter dürfte es sein, die Bewertung der Daten anzubieten und dann in der *Cloud* zu speichern und auszuwerten. Der Nachteil einer solchen Überlegung besteht darin, dass die Leser zum Ranking der Einträge motiviert werden müssen. Die vorgestellte Idee gewährleistet dies, indem sie den Bewertungsprozess zeitlich vor die Übernahme der Metadaten in das Literaturverwaltungsprogramm setzen würde.

Außer den Benutzern müssen auch die Betreiber von Repositorien dazu motiviert werden, OA_SCORE zu implementieren. Erste Voraussetzung ist hierfür, dass problemlos zu integrierende Plug-ins bereitgestellt werden müssen. In einem nächsten Schritt müsste mit dem sich ergebenden Mehrwert argumentiert werden. Für zentrale Repositorien könnte es beispielsweise erstrebenswert sein, möglichst viele gut bewertete Items veröffentlicht zu haben und damit eine Vorrangstellung innerhalb der Fachdisziplin zu behaupten. Für institutionelle Repositorien gilt im Prinzip das Gleiche. Solange aber Qualitätsstandards eine Bewertung nicht vorschreiben, haben Institutionen auf den unteren Rängen kaum einen Grund, ihre potentiell schlechteren Veröffentlichungen bewerten zu lassen. Ein Ansatz wäre folglich die Aufnahme von OA_SCORE in ein System von Qualitätsstandards. Das ist aber nur denkbar, wenn das Instrument zuvor seine Güte und Zuverlässigkeit unter Beweis gestellt hat.

Um Autoren unabhängig von Repositorienbetreibern die Bewertung ihrer Veröffentlichungen mit Hilfe von OA_SCORE zu ermöglichen, sollte - wie bereits erwähnt - eine Integration über Verlinkung angeboten werden. Dem Autor wird dazu ein Link zur Verfügung gestellt, über den die Leser an ein angereichertes Dokument gelangen. Ein Modell des Prozesses ist in **Abbildung 15** dargestellt.

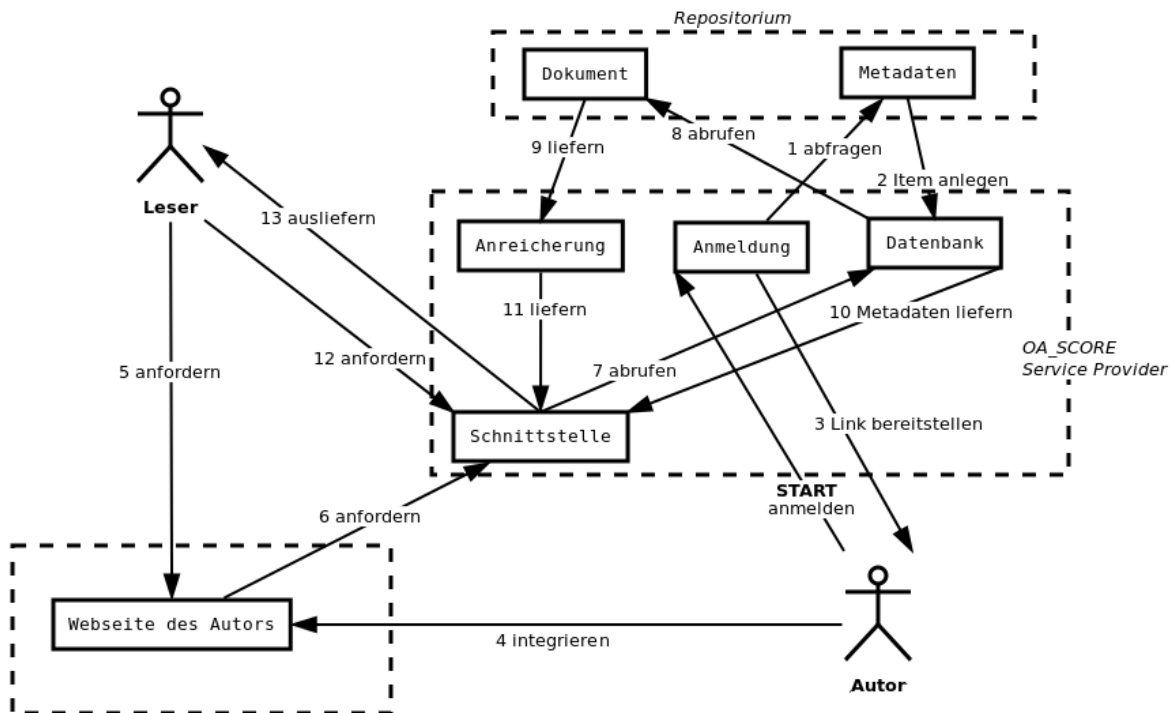


Abbildung 15: Kollaborationsdiagramm für OA_SCORE-Verlinkung

Zusätzlich hat der Autor die Möglichkeit mit Hilfe eines *Inlineframe* den aktuellen Stand der Bewertung anzeigen zu lassen. Dieses Verfahren hat für alle Beteiligten Vor- und Nachteile, stellt jedoch insgesamt gesehen eine Bereicherung des Konzepts dar (**Tabelle 26**).

Als Konsequenz dürften Betreiber von Repositorien zusätzlich motiviert sein, OA_SCORE ins Repository zu integrieren.

8 Schluss und Ausblick

Es ist eine Besonderheit der vorliegende Arbeit, dass sie ein derzeit noch nicht existierendes Werkzeug auf seine Anwendbarkeit hin untersucht. Angesichts des zur Verfügung gestellten Rahmens liegt es in der Natur der Sache, dass nicht alle Aspekte ausreichend beleuchtet werden konnten. Insbesondere konnte kein eindeutiger Nachweis erbracht werden, inwieweit das vorgestellte Werkzeug wirklich in der Lage ist, die Erwartungen zu erfüllen. Erwartet wird, dass OA_SCORE kurzfristig gültige und nachvollziehbare Aussagen hinsichtlich der inhaltlichen Qualität einer wissenschaftlichen Arbeit treffen kann. Es sollte sich dabei immer nur um eine Ergänzung zu anderen Verfahren, wie Peer Review, Zitationsanalyse, *Usage* und *Altmetrics* handeln.

Die von Charles Jennings aufgezeigten Hürden für jede neue Methode der Qualitätsmessung können von OA_SCORE mit teilweisen Einschränkungen überwunden werden.

Die Einschränkungen betreffen zum einen die nicht ausreichend geklärte Reliabilität von OA_SCORE. Die angewandte Methode ist angreifbar, da sie einerseits auf einen zu geringen Datenkorpus beruht und andererseits nicht ausreichend genau die Vergleichswerte justieren konnte.

Perspektive	Vorteile	Nachteile
Autor	Kann unabhängig vom Repositorium eine Bewertung anbieten. Verlinkung ist praktisch überall einsetzbar (Homepage, Social Media, usw.)	Muss selbst aktiv werden.
Repositorium	Integration von OA_SCORE ins Repositorium entfällt.	Herkunft der Leser kann nicht mehr direkt nachverfolgt werden. Datenintegrität muss sichergestellt werden. Keine generelle Bewertung aller Items möglich
OA_SCORE	Unabhängigkeit von Repositorien. Verringerter Programmieraufwand. Höhere Reichweite. Es werden auch Items erfasst, die noch nicht bewertet wurden. Zusätzliche Mehrwertdienste können angeboten werden.	Erhöhter Bedarf an Rechenleistung. Dokumentaushlieferung kann sich verzögern. Repositorien müssen automatischen Download von Dokumenten erlauben. Änderungen der Metadaten müssen separat erfasst werden.
Leser	Darstellung der Metadaten ist vereinheitlicht, Benutzerfreundlichkeit wird verbessert.	Muss zweimal klicken um das Dokument anzufordern.

Tabelle 26: Vor- und Nachteile der Integration von OA_SCORE durch Verlinkung

Eine weitere Einschränkung ist der unvollständige Schutz vor Manipulationen. Sie können zwar in ihrer Zahl eingeschränkt, jedoch nicht gänzlich ausgeschlossen werden. Durch die Einbindung von ORCID kann die Wahrscheinlichkeit von Manipulationen jedoch stark reduziert werden.

Ein besonderes Problem stellt die Tatsache dar, dass die zu bewertenden Kriterien nicht definiert werden konnten. Es konnte lediglich gezeigt werden, dass die von den erfolgreichen Zeitschriften verwendeten Kriterien, insbesondere die der Originalität und der Relevanz, nicht ohne weiteres auf OA_SCORE übertragbar sind. Eventuell stellt die innerhalb des *Beyond-Impact*-Projekts gestartete Diskussion um geeignete Indikatoren Lösungen für dieses Problem in Aussicht. Während des ersten Workshops wurden dort die Indikatoren *position*, *productivity*, *quality*, *relevance*, *leadership*, *translation*, *innovation*, *economic impact* und *societal impact* identifiziert (**Beyond Impact 2011**). Leider scheint aber die Diskussion innerhalb der Arbeitsgruppe seit 2011 nicht weiter vorangekommen zu sein.

Alle genannten Einschränkungen stellen aber keinen Grund dafür dar, OA_SCORE generell abzulehnen. Vielmehr entstehen dadurch Ideen für zukünftige Forschungsfragen.

Auf der Habenseite steht der Nachweis, dass PDF das mit Abstand häufigste Dateiformat zur Veröffentlichung von aktuellen wissenschaftlichen Ergebnissen im OA-Bereich darstellt. Somit steht der Grundidee von OA_SCORE, nämlich der zeitgleichen Bewertung von mehreren Kriterien im zu bewertenden Dokument selbst prinzipiell nichts im Wege.

9 Literaturverzeichnis

9.1 Institutionelle URL

Academic Journals. Zugegriffen 20. Mai 2013. <http://academicjournals.org/> .

Bielfeld Academic Search Engine. Zugegriffen 23. Mai 2013. <http://base-search.net/>

Copernicus Publications. *Public Peer-Review*. Zugegriffen 12. Mai 2013. http://publications.copernicus.org/services/public_peer_review.html/ .

Cybermetrics Lab. *Ranking Web of Universities*. Zugegriffen 20. Mai 2013. <http://www.webometrics.info/>.

DuraSpace. *Dspace*. Zugegriffen 20. Mai 2013. <http://www.dspace.org/> .

F1000 Research Ltd. *Faculty of 1000*. Zugegriffen 20. Mai 2013. <http://f1000.com/> .

FORCE11. Zugegriffen 20. Mai 2013. <http://www.force11.org/> .

Google. *Google Ngram Viewer*. Zugegriffen Februar 17, 2013. <http://books.google.com/ngrams/info>.

ImpactStory. Zugegriffen 3. Dezember 2013. <http://impactstory.org/> .

The ISNI Organization. Zugegriffen 20. Mai 2013. <http://www.isni.org/> .

PSNC Digital Libraries Team. *Strona domowa systemów dArceo, dLab, dLibra oraz dMuseion*. Zugegriffen 20. April 2013. <http://dlab.psnc.pl/> .

ORCID Inc. Zugegriffen 12. Mai 2013. <http://orcid.org/> .

Universitätsbibliothek Regensburg. *Elektronische Zeitschriftenbibliothek*. Zugegriffen 23. Mai 2013. <http://ezb.uni-regensburg.de/> .

University of Nottingham. *OpenDOAR - Directory of Open Access Repositories*. Zugegriffen 20. Mai 2013. <http://www.opendoar.org/> .

University of Southampton. *EPrints - Digital Repository Software*. Zugegriffen 20. Mai 2013. <http://www.eprints.org/> .

9.2 Publikationen

Adobe Systems Incorporated. „PDF Reference and Adobe Extensions to the PDF Specification.“ Zugegriffen 20. Mai 2013. http://www.adobe.com/devnet/pdf/pdf_reference.html .

Adobe Systems Incorporated. „Document management — Portable document format —Part 1: PDF 1.7“, 2008, http://www.adobe.com/content/dam/Adobe/en/devnet/acrobat/pdfs/PDF32000_2008.pdf

American Association for the Advancement of Science (ed.). „Science’s Instructions for Reviewers“. Zugegriffen Februar 17, 2013. <http://www.sciencemag.org/site/feature/contribinfo/refinstr.xhtml> .

Andreasson, Stefan. „Orientalism and African Development Studies: the ‘reductive repetition’ motif in theories of African underdevelopment“. *Third World Quarterly* 26, Nr. 6 (2005): 971–986. doi:10.1080/01436590500089307 .

Arnold, Douglas N., und Kristine K. Fowler. „Nefarious Numbers.“ *arXiv:1010.0278* (Oktober 1, 2010). <http://arxiv.org/abs/1010.0278> .

Bealls, J. „Predatory Publishing Update“. Zugegriffen Februar 23, 2013. <http://scholarlyoa.com/2012/05/24/predatory-publishing-update/> .

Bellini, E., M. A. Deussom, und P. Nesi. „Assessing open archive oai-pmh implementations.“ In *The 16th International Conference on Distributed Multimedia Systems*, 2010. <http://www.disit.dsi.unifi.it/articoli/DMS2010-OAI-PMH-final-v9-1.pdf> .

FORCE11 (ed.) „Beyond Impact Workshop Report“. Zugegriffen 12. Mai 2013. https://docs.google.com/document/d/1sH3JOW5Luki4i37Ve1mOnI2wNZJ-baUOx1T42S_7txQ0/edit?hl=en_GB&pli=1#heading=h.jzliwyb28tr .

Björk, Bo-Christer, und David Solomon. „Open Access Versus Subscription Journals: a Comparison of Scientific Impact“. *BMC Medicine* 10, Nr. 1 (Juli 17, 2012): 73. doi:10.1186/1741-7015-10-73.

Borenstein, Nathaniel S., und Ned Freed. „Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies“. 1996. Zugegriffen 21. Mai 2013. <http://tools.ietf.org/html/rfc2045> .

Borlund, Pia. „The concept of relevance in IR“. *Journal of the American Society for information Science and Technology* 54, Nr. 10 (2003): 913–925. doi: 10.1002/asi.10286

Cell. „Cell - Reviewers“. Zugegriffen Februar 17, 2013. <http://www.cell.com/Reviewers/> .

Clarke, Maxine. „Report of Nature’s peer review trial“, December 20, 2006, http://blogs.nature.com/peer-to-peer/2006/12/report_of_natures_peer_review_trial.html

Dago (Pseudonym): „XMLStarlet Command Line XML Toolkit: Overview“. Zugegriffen 20. Mai 2013. <http://xmlstar.sourceforge.net/overview.php> .

Deutsche Initiative für Netzwerkinformation e. V. „DINI-Zertifikat Dokumenten- und Publikationsservice“. 2010.2010

Digital Library Federation (ed.) „METS: Metadata Encoding and Transmission Standard“ . Zugegriffen 25. April 2013. <http://www.loc.gov/standards/mets/mets.xsd>.

Directory of Open Access Journals. „DOAJ hits the 1 MILLION mark!“, 2013-03-12 Zugegriffen 20. Mai 2013. <http://www.doaj.org/doaj?func=news&nId=44&uiLanguage=en> .

Drümmer, Olaf, und Thomas Zellmann. „Langzeitarchivierung mit PDF/A-2“. *Informatik-Spektrum* 34, Nr. 5 (2011): 496–500. doi:10.1007/s00287-011-0560-9

Dublin Core Metadata Initiative (ed.). „Dublin Core Metadata Element Set, Version 1.1“. Zugegriffen Februar 16, 2013. <http://dublincore.org/documents/dces> .

Force11 (ed.). „Beyond the PDF2: Outcomes “. Zugegriffen 20. Mai 2013. <http://www.force11.org/outcomes>.

Google Scholar. „Top-Publikationen - Englisch“. Zugegriffen 20. Mai 2013. http://scholar.google.com/citations?view_op=top_venues&hl=de .

Haider, Jutta. „Open Access hinter verschlossenen Türen oder wie sich Open Access im und mit dem Entwicklungsdiskurs arrangiert “ in: *„Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft“*, 65-84 (3. Mai 2012). **urn:nbn:de:bsz:291-universaar-873**

Henneberger, Sabine. „Entwicklung einer Analysemethode für Institutional Repositories unter Verwendung von Nutzungsdaten“; Dissertation, Humboldt-Universität zu Berlin, Philosophische Fakultät I, publiziert am 31.10.2011, **urn:nbn:de:kobv:11-100196484**

Herb, Ulrich. „Offenheit und wissenschaftliche Werke: Open Access, Open Review, Open Metrics, Open Science & Open Knowledge “ in: *„Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft“*, 11-44 (3. Mai 2012). **urn:nbn:de:bsz:291-universaar-873**

Herb, Ulrich. „OpenAccess Statistics: alternative impact measures for Open Access documents? : an examination how to generate interoperable usage information from distributed Open Access services“ (September 28, 2010). **urn:nbn:de:bsz:291-scidok-33766**

Internet Assigned Numbers Authority (ed.) „MIME Media Types“. Zugegriffen 20. Mai 2013. **<http://www.iana.org/assignments/media-types>**.

Jennings, Charles. „Quality and value: The true purpose of peer review“. *Nature* (2006). **doi:10.1038/nature05032**.

Jovanović, Miloš. „Eine kleine Frühgeschichte der Bibliometrie“. *Information - Wissenschaft & Praxis* 63, Nr. 2 (2012): 71-80. **doi:10.1515/iwp-2012-0017**

Lane, Julia. „Let’s make science metrics more scientific“. *Nature* 464, Nr. 7288 (2010): 488-489. **doi:10.1038/464488a**

The Lancet. „TheLancet.com“. Zugegriffen Februar 17, 2013. **<http://www.thelancet.com/protocol-reviews>**

Lagoze, Carl, Herbert Van de Sompel, Michael Nelson, und Simeon Warner. „Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0“, 14. Juni 2002. **<http://www.openarchives.org/OAI/openarchivesprotocol.html>** .

LeCun, Yann. *DjVu: A Tutorial*. Zugegriffen 20. April 2013. **<http://www.djvuzone.org/support/tutorial/chapter-serving.html>** .

Nature (ed). „Peer-review policy : authors & referees @ npg“. Zugegriffen Februar 17, 2013. **http://www.nature.com/authors/policies/peer_review.html** .

Nelson, M. L., J. A. Smith, und I. G. Del Campo. „Efficient, automatic web resource harvesting“. In *Proceedings of the 8th annual ACM international workshop on Web information and data management*, 43-50, 2006. **doi:10.1145/1183550.1183560**

The New England Journal of Medicine. „NEJM - Author Center“. Zugegriffen Februar 17, 2013. **<https://cdf.nejm.org/misc/authors/>**.

Nottingham, Mark, Dave Crocker, und Peter Saint-Andre. „Deprecating the ‚X-‘ Prefix and Similar Constructs in Application Protocols“. 2012. Zugegriffen 21. Mai 2013. **<http://tools.ietf.org/html/rfc6648>** .

Oppenheim, Charles. „Electronic Scholarly Publishing and Open Access“. *Journal of Information Science* 34, Nr. 4 (Januar 8, 2008): 577-590. **doi:10.1177/0165551508092268**.

Priem, Jason, Carolina-Chapel Hill, Dario Taraborelli, Paul Groth, und Cameron Neylon. „altmetrics: a manifesto – altmetrics.org“. 2010. Zugegriffen Januar 31, 2013. **<http://altmetrics.org/manifesto/>** .

„RFC 2046 - Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types“. Zugegriffen 20. April 2013. <http://tools.ietf.org/html/rfc2046>.

Shotton, David. „The Five Stars of Online Journal Articles - a Framework for Article Evaluation“. *D-Lib Magazine* 18, Nr. 1/2 (Januar 2012). doi:10.1045/january2012-shotton.

Taraborelli, D. „Soft peer review: social software and distributed scientific evaluation“ Proceedings paper, 2008. <http://coop.wineme.fb5.uni-siegen.de/?id=coop2008>.

Willmann, Urs. „Kontinentalverschiebung: ‚Kein schrumpfender Apfel‘“. *Die Zeit*, 6. Januar 2012, Abschn. Umwelt. <http://www.zeit.de/2012/02/G-Plattentektonik> .

Van de Sompel, H., M. L. Nelson, C. Lagoze, und S. Warner. „Resource harvesting within the OAI-PMH framework“. *D-lib magazine* 10, Nr. 12 (2004): 1082-9873. doi:10.1045/december2004-vandesompel

Van Ossenbruggen, Jacco. „Processing Structured Hypermedia—A Matter of Style“. PhD thesis, Vrije Universiteit, Amsterdam, The Netherlands, 2001. <http://homepages.cwi.nl/~jrvosse/thesis/> .

Vierkant, Paul. „Visualizing Open Access“. *LIBREAS. Library Ideas* 8, Nr. 21 (2012). urn:nbn:de:kobv:11-100204968

Vullo, G., G. Clavel, N. Ferro, S. Higgins, R. van Horik, W. Horstmann, und S. Kapidakis. „Quality interoperability within digital libraries: the DL.org perspective“. In *Pre-proceedings of 2nd DL.org Workshop*, 12-24. Glasgow, UK, 2010. <http://eprints.gla.ac.uk/46123/> .

10 Anhang

10.1 Häufigkeits- und Synonymanalyse der Begriffsgruppen Originalität, Plausibilität und Nützlichkeit

Datei: Jung_OA_SCORE__synonymanalyse.ods

Datei: Jung_OA_SCORE__synonymanalyse.pdf

10.2 Rohdaten der Meta-Repositorien

10.2.1 OpenDOAR

Die komplette OpenDOAR-Datenbank kann über die API-Schnittstelle <http://www.open-doar.org/api13.php?all=y> abgerufen werden. Für die Untersuchung wurde am 26.1.2013 ein Auszug erstellt.

Datei: Jung_OA_SCORE__raw_OpenDOAR_130126.xml

10.2.2 ROAR

Beim *Registry of Open Access Repositories* (ROAR) kann der jeweils aktuelle Datenbestand über ein Formular auf der Startseite <http://roar.eprints.org> exportiert werden. Für die Untersuchung wurde das Schema „EP 3 XML“ verwendet. Der Datenauszug wurde am 22.1.2013 heruntergeladen.

Datei: Jung_OA_SCORE__raw_ROAR_130122.xml

10.3 Konsolenbefehle zur Abfrage der Meta-Repositorien mittels *xmlstart*

Über eine Gruppe von BASH-Befehlen lassen sich die gewünschten Daten aus den beiden Datenbank-Auszügen extrahieren. Das Tool *xmlstarlet* wird vorausgesetzt.

Datei: Jung_OA_SCORE__konsole

Datei: Jung_OA_SCORE__konsole.pdf

10.4 BASH-Script zur Datenabfrage nach OAI-PMH

Mittels dieses BASH-Scripts wurden Daten aus den durch Schnittstellenadresse zugänglichen Repositorien mittels OAI-PMH abgefragt. Die Datei mit den Schnittstellenadressen *oaiurl* muss sich im gleichen Verzeichnis befinden. Das Tool *xmlstarlet* muss installiert sein.

Datei: Jung_OA_SCORE__script.sh

Datei: Jung_OA_SCORE__script.pdf

10.5 Verwendete Fragebögen

Pro Sprache gibt es einen Fragebogen. Außerdem wurden den Befragten, nachdem sie die 2. Frage beantwortet hatten, ein Blatt mit den vorbereiteten Skalen und den damit verbundenen Hinweisen vorgelegt. Die vorbereiteten Etiketten mit den Achsenbezeichnungen wur-

den dazu mit Hilfe von doppelseitigem Klebeband angebracht und die beiden Blätter mittels einer transparenten Prospekthülle zu einer Einheit zusammengefasst.

Datei: Jung_OA_SCORE_Fragebogen_de.pdf

Datei: Jung_OA_SCORE_Fragebogen_de_Beispielskala_O_P_R_P_R.pdf

Datei: Jung_OA_SCORE_Fragebogen_fr.pdf

Datei: Jung_OA_SCORE_Fragebogen_fr_Beispielskala_O_F_V_F_V.pdf

10.6 Korrigierte Rohdaten der Repositorien

Die Daten liegen in zwei verschiedenen Dateiformaten vor: *text/csv* und *application/xml* (Gnumeric Default XML-based File). Die GNUMERIC-Datei wurde um Spaltensummen ergänzt.

Datei: Jung_OA_SCORE_Daten_Harvesting.csv

Datei: Jung_OA_SCORE_Daten_Harvesting.gnumeric

10.7 Analyse der Metadatenformate

Es wurden die Angaben zum *metadataPrefix* aus den korrigierten Rohdaten entnommen und in Bezug auf die Häufigkeit ihres Auftretens analysiert.

Datei: Jung_OA_SCORE_Analyse_Metadatenformate.ods

10.8 Untersuchung zur Auftreten von Formatangaben

Datei: Jung_OA_SCORE_Untersuchung_der_Formatangaben.ods

Datei: Jung_OA_SCORE_Untersuchung_der_Formatangaben.pdf

10.9 Überprüfung zur Zahl der begutachteten Items

Datei: Jung_OA_SCORE_Stichproben_zur_Zahl_der_Begutachtung.pdf

10.10 Ergebnisse der Befragung

Die Dateien enthalten sowohl die Rohdaten, als auch die operationalisierten Daten aus der Befragung.

Hinweis: Das LibreOffice-Dokument enthält zwei Makros. Für den Fall, dass diese nicht aktiviert werden sollen, steht zusätzlich ein PDF-Dokument zur Verfügung.

Datei: Jung_OA_SCORE_Ergebnisse_Befragung.ods

Datei: Jung_OA_SCORE_Ergebnisse_Befragung.pdf

10.11 XML-Schema

Das vorgestellte XML-Schema basiert auf einem Entity-Relationship-Modell für den Datenaustausch von OA_SCORE.

Datei: Jung_OA_SCORE_ER-Modell_für_Datenaustausch.png

Datei: Jung_OA_SCORE_xml-Schema.xsd

Datei: Jung_OA_SCORE_xml-Beispiel.xml

Abbildungsverzeichnis

Abbildung 1: ER-Modell von OA_SCORE.....	13
Abbildung 2: Beispiel einer zweidimensionalen Skala (2D-Skala).....	21
Abbildung 3: Entwicklungsschritte zu einer 3D-Skala.....	21
Abbildung 4: Beispiel einer dreidimensionalen Skala (3D-Skala).....	22
Abbildung 5: Beispiel einer 1-von-8-Skala.....	22
Abbildung 6: Anteil der Repositorien (n=MF=1036), bei der die Gesamtzahl der abgefragten Records ein Vielfaches von bestimmten Werten ist.....	31
Abbildung 7: Häufigkeitsverteilung von Metadatenformaten in den untersuchten Repositorien (abgebildet sind nur die am häufigsten vorkommenden Formate).....	33
Abbildung 9: Verteilung des Verhältnisses der Anzahl von Records mit Formatangaben und der Anzahl der Formatangaben selbst.....	35
Abbildung 8: Verteilung des Verhältnisses der Formatangaben je Record).....	34
Abbildung 10: Verteilung der gefundenen Formatangaben	37
Abbildung 11: Anteil des MIME-Typs application/pdf an der Gesamtzahl der Formatangaben nach zusammengefassten Häufigkeitsklassen.....	38
Abbildung 12: Erläuterung zur Analyse der Werte aus der 3D-Skala.	45
Abbildung 13: Beispiel einer verbesserten verweissensitiven 1-von-8-Skala	49
Abbildung 14: Kollaborationsdiagramm des OA_SCORE Service Provider	52
Abbildung 15: Kollaborationsdiagramm für OA_SCORE-Verlinkung.....	54

Tabellenverzeichnis

Tabelle 1: Übersicht zu den derzeitigen Filterklassen.....	12
Tabelle 2: Übersicht über die Suchpattern zur Ermittlung von begutachteten Veröffentlichungen.....	19
Tabelle 3: Zusammengestellte Bewertungskriterien für Reviewers der fünf bestplatzierten Journals nach GSM.....	23
Tabelle 4: Ermittelte Begriffe und Synonyme für die zu bewertenden Kategorien in vier Sprachen.....	26
Tabelle 5: Charakteristik der zur Befragung verwendeten Texte.....	29
Tabelle 6: Übersicht über die Menge der untersuchten Repositorien	30
Tabelle 7: Repositorien, deren Formatangaben möglicherweise mittels RDF bzw. DIDL identifiziert werden könnten.....	34
Tabelle 8: Übersicht über die Verteilung von Formatangaben.....	34
Tabelle 9: Anteil der MIME-Typen „text/plain“ in DSpace-Repositorien im Vergleich zur Gesamtzahl dieses MIME-Typ (Hinweis: Es wurden nur solche Repositorien gezählt, die das Wort „dspace“ in der URL führen.).....	38
Tabelle 10: Verteilung des MIME-Typs application/pdf bezogen auf den Typ des Items.....	39
Tabelle 11: Verteilung des MIME-Typs application/pdf bezogen auf den Typ des Records und Repositorien, in denen mindesten 5 % begutachtete MIME-Typen application/pdf gezählt wurden.....	39
Tabelle 12: Ergebnis der Kontrollzählung von 14 Repositorien in Bezug auf die Verteilung von begutachteten PDF-Dokumenten.....	40
Tabelle 13: Verteilung der Vorlieben für bestimmte Kategoriebezeichnungen.....	42
Tabelle 14: Häufigkeiten der Achsen-Definitionen bei der 2D-Skala (Paare).....	42
Tabelle 15: Häufigkeiten der Achsen-Definitionen bei der 2D-Skala (Einzelne Kriterien).....	42
Tabelle 16: Arithmetische Mittel der einzelnen Kategorien aus der 2D-Skala für die untersuchten Texte.....	43
Tabelle 17: Median der einzelnen Kategorien aus der 2D-Skala für die untersuchten Texte	44
Tabelle 18: Arithmetisches Mittel der einzelnen Kategorien aus der 2D-Skala für die untersuchten Texte, jeweils auf ganze Zahlen gerundet (vergl. Tabelle 17).	44
Tabelle 19: Arithmetische Mittel der einzelnen Kategorien aus der 3D-Skala für die untersuchten Texte.....	45
Tabelle 20: Arithmetische Mittel der gerichteten extrahierten Abstandswerte.....	46
Tabelle 21: Zusammengefasste arithmetische Mittel der gerichteten extrahierten Abstandswerte.....	46
Tabelle 22: Arithmetische Mittel der aus der 1-von-8-Skala extrahierten Werte	46
Tabelle 23: Untermedian der aus der 1-von-8-Skala extrahierten Werte (n: 27 - 29).....	47
Tabelle 24: Von den Befragten bevorzugte Skalen (n=27).....	47
Tabelle 25: Vergleich zwischen 2D-Skala und 1-von-8-Skala.....	48
Tabelle 26: Vor- und Nachteile der Integration von OA_SCORE durch Verlinkung.....	55